

M.ª Ángeles Cea D'Ancona



Fundamentos y aplicaciones en metodología cuantitativa

FUNDAMENTOS Y APLICACIONES EN METODOLOGÍA CUANTITATIVA

Consulte nuestra página web: www.sintesis.com
En ella encontrará el catálogo completo y comentado

FUNDAMENTOS Y APLICACIONES EN METODOLOGÍA CUANTITATIVA

M.^a Ángeles Cea D'Ancona



EDITORIAL
SÍNTESIS

Reservados todos los derechos. Está prohibido, bajo las sanciones penales y el resarcimiento civil previstos en las leyes, reproducir, registrar o transmitir esta publicación, íntegra o parcialmente, por cualquier sistema de recuperación y por cualquier medio, sea mecánico, electrónico, magnético, electroóptico, por fotocopia o por cualquier otro, sin la autorización previa por escrito de Editorial Síntesis, S. A.

© M.^a Ángeles Cea D'Ancona

© EDITORIAL SÍNTESIS, S. A.
Vallehermoso, 34 - 28015 Madrid
Teléf.: 91 593 20 98
[http: //www.sintesis.com](http://www.sintesis.com)

Depósito Legal: M. 25.149-2012
ISBN: 978-84-9958-905-3

Impreso en España - Printed in Spain

*A mis maestros y lectores,
con gratitud*

ÍNDICE

INTRODUCCIÓN	11
 CAPÍTULO 1: LA METODOLOGÍA CUANTITATIVA Y EL ANÁLISIS DE LA REALIDAD SOCIAL	 13
1.1. Fundamentos de la metodología cuantitativa en la investigación social	13
1.1.1. Antecedentes de la investigación social empírica	14
1.1.2. La fundamentación teórica: del positivismo al funcionalismo	17
1.2. Características distintivas de la metodología cuantitativa <i>versus</i> cualitativa	22
1.3. La articulación metodológica	29
1.4. Del método científico al proyecto de investigación	35
1.4.1. Integrantes esenciales del método científico	37
1.4.2. Distinción entre proyecto y diseño de investigación	43
1.4.3. Tipologías de diseños de investigación	51
1.5. La calidad de la investigación social	62
Lecturas complementarias	65
Ejercicios propuestos	65
 CAPÍTULO 2: LA MEDICIÓN DE LA REALIDAD SOCIAL	 67
2.1. Fundamentos de la medición	67
2.2. El proceso de medición: de los conceptos teóricos a los indicadores empíricos	 69
2.3. Tipologías de variables	88
2.4. Escalas de medición de actitudes	92
2.4.1. Escala de distancia social de Bogardus	92
2.4.2. Escala diferencial de Thurstone	95
2.4.3. Escalograma o escala acumulativa de Guttman	97

2.4.4. Escala aditiva de Likert	101
2.4.5. Diferencial semántico de Osgood	108
2.5. La calidad de la medición: fiabilidad y validez	111
2.5.1. Validez de la medición	112
2.5.2. Comprobación de la fiabilidad	117
Lecturas complementarias	125
Ejercicios propuestos	126
 CAPÍTULO 3: LA LÓGICA EXPERIMENTAL EN LA INVESTIGACIÓN SOCIAL	
3.1. Fundamentos de la experimentación	127
3.2. Características esenciales de la experimentación	128
3.3. Pluralidad de diseños experimentales	136
3.4. Amenazas a la validez del experimento	149
3.5. Ventajas e inconvenientes de la experimentación como estrategia de investigación	152
Lecturas complementarias	154
Ejercicios propuestos	154
 CAPÍTULO 4: LA INVESTIGACIÓN “NO REACTIVA” A PARTIR DE DATOS DISPONIBLES	
4.1. La medición “no reactiva”	155
4.2. Investigar a partir de <i>fuentes secundarias</i> : archivos bibliográficos, documentales, estadísticos, de prensa	157
4.2.1. Revisión de investigaciones teórico-empíricas publicadas	158
4.2.2. Análisis de datos secundarios	160
4.2.3. Meta-análisis	169
4.2.4. Análisis de contenido cuantitativo	174
4.3. Evaluación de los datos secundarios	180
4.4. Ventajas e inconvenientes de la investigación a partir de datos disponibles	181
Lecturas complementarias	183
Ejercicios propuestos	184
 CAPÍTULO 5: LA ENCUESTA COMO ESTRATEGIA DE INVESTIGACIÓN SOCIAL	
5.1. Fundamentos de la encuesta	185
5.1.1. Configuración de la encuesta “por muestreo”	189
5.1.2. Dominio de la encuesta por muestreo en la investigación social norteamericana	191
5.1.3. Desarrollo de la investigación mediante encuesta en España	196
5.2. Características esenciales de la encuesta y su articulación con otras estrategias de investigación	198
5.2.1. Posibilidades y límites de la encuesta	199

5.2.2. La articulación de la encuesta con otras estrategias de investigación	201
5.3. Fases en el desarrollo de la encuesta y errores en su aplicación	206
5.3.1. Tipología de errores de encuesta	209
5.4. Modalidades de encuesta	212
5.4.1. Encuesta “cara a cara” o mediante entrevista presencial	215
5.4.2. Encuesta telefónica	221
5.4.3. Encuestas autoadministradas	225
5.4.4. Métodos mixtos de encuesta	230
Lecturas complementarias	232
Ejercicios propuestos	233
 CAPÍTULO 6: EL DISEÑO Y LA EVALUACIÓN DE CUESTIONARIOS	235
6.1. Elaboración del cuestionario	235
6.1.1. Actuaciones previas a la redacción de preguntas	237
6.1.2. La redacción de las preguntas	240
6.1.3. El formato de la pregunta	246
6.1.4. La ubicación de la pregunta en el cuestionario	258
6.1.5. La forma del cuestionario	261
6.1.6. Tratamiento de temas sensibles y referidos al pasado	267
6.2. Evaluación del cuestionario	274
Lecturas complementarias	280
Ejercicios propuestos	281
 CAPÍTULO 7: LA SELECCIÓN DE UNIDADES DE OBSERVACIÓN: DISEÑO MUESTRAL Y TRABAJO DE CAMPO	283
7.1. Fundamentos de muestreo	283
7.2. Decisiones claves en el diseño muestral	285
7.2.1. Delimitación de la población de estudio	286
7.2.2. Elección del <i>marco muestral</i> y su traducción en <i>errores de cobertura</i>	288
7.2.3. Determinación del tamaño de la muestra	293
7.3. Selección de las unidades de la muestra: muestreos probabilísticos y no probabilísticos	298
7.3.1. Muestreo aleatorio simple	299
7.3.2. Muestreo aleatorio sistemático	303
7.3.3. Muestreo aleatorio estratificado	304
7.3.4. Muestreo aleatorio por conglomerados	311
7.3.5. Muestreo de áreas y rutas aleatorias	320
7.3.6. Métodos de selección de individuos en una misma vivienda	328
7.3.7. Diseños muestrales no probabilísticos	333
7.4. Determinantes del error muestral	336
7.5. El error de “no respuesta” y su reducción	344
Lecturas complementarias	353
Ejercicios propuestos	353
Anexo: Tabla de número aleatorios	356

CAPÍTULO 8: ANÁLISIS DE DATOS Y PRESENTACIÓN DE RESULTADOS	357
8.1. Preparación de los datos para el análisis	357
8.2. Análisis exploratorio univariable	362
8.3. Análisis cruzado bivariable	368
8.4. Análisis multivariable	374
8.5. Informe de la investigación: contenidos básicos y redacción	383
Lecturas complementarias	385
Ejercicios propuestos	386
 BIBLIOGRAFÍA CITADA	 391

INTRODUCCIÓN

Desde que en 1989 comenzara mi carrera docente, y en especial tras la publicación de mi primer libro sobre *metodología cuantitativa* en 1996, la andadura a través de los *métodos y técnicas* de investigación social ha ido progresando. Se ha beneficiado de lo aportado por otros autores e investigadores, de reflexiones colectivas con colegas y alumnos, y de la experiencia propia a través de la investigación y la docencia. El libro que el lector tiene en sus manos pretende reflejar el camino andado, a la vez que seguir contribuyendo al interés y conocimiento de la *metodología cuantitativa*. Quiere ofrecer una visión panorámica y pormenorizada de cada uno de sus componentes principales, de sus *fundamentos, aplicaciones e innovaciones* que reflejan su incesante avance.

El libro se estructura en ocho capítulos. El primero versa sobre la distintividad de la *metodología cuantitativa* y los puntos de encuentro con la metodología cualitativa, en tanto vías complementarias de acceso al análisis de la realidad social. Se parte de la pluralidad del *objeto* y del *método*, de la necesidad de articulación o convergencia interparadigmática ante la complejidad de los fenómenos sociales. El capítulo comienza, al igual que los siguientes, por los *fundamentos* que desde la *teoría* y la *empiría* han contribuido a la configuración de la *metodología cuantitativa* en la investigación social. Se hablará del *método científico* y su traducción en un *proyecto* de investigación. También de *calidad*, un integrante básico cada vez más exigido en la investigación para que sus resultados sean creíbles y reviertan en el conocimiento de la realidad social.

El capítulo segundo aborda la *medición* de la realidad social, sus fundamentos, la traducción de *conceptos* teóricos a *indicadores* empíricos y su síntesis en *índices*. Atención específica tienen las escalas de medición de las actitudes, al igual que los criterios de *calidad* de la medición; las diferentes modalidades de *validez* y de *fiabilidad*, junto con las actuaciones para su comprobación y consecución. A este capítulo siguen

los específicos a las tres estrategias fundamentales en la metodología cuantitativa: la *experimentación*, la investigación *no reactiva* y la *encuesta*. En todos ellos se hablará de fundamentos, características diferenciales, sus diferentes modalidades, potencialidades y límites, junto con las actuaciones a seguir para su correcta realización.

Los tres últimos capítulos se dedican al diseño y evaluación de *cuestionarios*, la selección de las unidades de observación (el *diseño muestral* y el *trabajo de campo*), el *análisis de datos* y la *presentación* de los resultados. Su contenido queda descrito en el *Índice*, donde se señalan los apartados y subapartados que componen cada capítulo. En dicho sumario se ha podido observar que la extensión de los capítulos varía. Guarda relación con la complejidad de la temática que abordan y su protagonismo en la investigación social.

Para ayudar a la comprensión de los contenidos teóricos, éstos se acompañan de ejemplos ilustrativos y de referencias a estudios teóricos y empíricos. Como podrá observarse, todos los capítulos incluyen *ejemplos* tomados de investigaciones “*reales*”, que sirven para ilustrar el *ideal* metodológico y el *hacer* concreto, la *teoría* y la *práctica* de la investigación social. Investigaciones publicadas en fechas próximas, de acceso a través de la Red, y que nos aproximan a la realidad de la investigación en contextos cercanos. Facilitan la comprensión del ejercicio de investigación, además de ofrecer referencias que pueden ser de utilidad para la actividad profesional.

Todos los capítulos concluyen con una selección de referencias bibliográficas, a modo de lecturas complementarias, y una propuesta de ejercicios. Con su realización puede comprobarse si realmente se han asimilado los contenidos esenciales del capítulo, sus conceptos teóricos y los desarrollos prácticos.

Pero antes de adentrarnos en la lectura de los capítulos, las últimas palabras de esta breve introducción son para expresar mi *gratitud* a mi familia, por su paciencia ante las privaciones que la redacción del presente libro ha supuesto. Vaya a ellos mi más profundo agradecimiento. También a los profesores y autores que han contribuido a mi formación, la mayoría de cuyos nombres aparecen listados en la bibliografía citada al final del libro. El *agradecimiento* se hace extensivo a los lectores, que hacen posible que se siga publicando, y a los editores, que de nuevo me han brindado la oportunidad de publicar y continuar contribuyendo al círculo del conocimiento.

1

LA METODOLOGÍA CUANTITATIVA Y EL ANÁLISIS DE LA REALIDAD SOCIAL

En 1985, Miguel Beltrán diferenciaba cinco vías de acceso al análisis de la realidad social: el método *histórico*, el *comparativo*, el *crítico-racional*, el *cuantitativo* y el *cualitativo*; dentro de una concepción pluralista tanto del *método* como del *objeto* de la sociología. A lo largo del presente capítulo se quiere ofrecer una visión panorámica de la metodología *cuantitativa* en la investigación social, cuyos componentes se irán desmenuzando en cada uno de los siete capítulos que van a continuación. En todos ellos la andadura comienza por los *fundamentos*, por las diferentes aportaciones que desde la *teoría* y la *investigación* han contribuido a su configuración. Como ya señalase Rodríguez Ibáñez (1989: 22), “si la historia aparece es porque la discusión acerca de la teoría y del método no puede prescindir de ella”. El paso siguiente es la descripción de sus características esenciales. En este caso, la distintividad de la metodología *cuantitativa versus* *cualitativa*, sus divergencias y puntos de encuentro. Debate al que sigue la defensa de su complementariedad, la articulación metodológica (o *triangulación*). Se describen las fases y los integrantes del *método científico* y cómo se traduce en un *proyecto* de investigación. El *proyecto* se distingue del *diseño* y las *estrategias* de las *técnicas* de investigación. El recorrido concluye con criterios de *calidad* que ha de cumplir el análisis de la realidad social que aplique metodología *cuantitativa*.

1.1. Fundamentos de la metodología cuantitativa en la investigación social

Aunque en apartados específicos de los capítulos correspondientes a estrategias principales de investigación, como la *experimentación*, la investigación *no reactiva* o la *encuesta*, se detallan las aportaciones que han contribuido a su fundamentación, en este apartado se ofrece una selección de aquellas que han contribuido a la cimentación de

la metodología *cuantitativa*. Aportaciones tanto desde el ángulo de la *teoría*, como de la *empiría*, teniendo como referente al método *científico* y como meta la cientificidad de la nueva ciencia: la “sociología”, término que acuña Auguste Comte en 1839, en el IV volumen de su *Curso de filosofía positiva* (*Cours de philosophie positive*), donde se resumen los principios básicos del *positivismo* (apartado 1.1.2).

1.1.1. Antecedentes de la investigación social empírica

Los antecedentes se remontan a los siglos XVII y XVIII, al movimiento denominado *Estadística Social*. Lo integran estudiosos (de ámbitos diferentes de conocimiento) a quienes une el interés por la recogida y organización de datos económicos, demográficos y sociales de la sociedad de su época. Promovieron que el estudio de los fenómenos sociales se hiciera siguiendo igual proceder que en las ciencias naturales. Su aportación se concreta en la aplicación de los avances realizados en el área de la estadística a la descripción de la sociedad. Incluye dos escuelas principales:

- Los *aritméticos políticos ingleses*: científicos ingleses que defienden que la cuantificación de las “regularidades” sociales se hiciera al modo de las naturales.

En 1660, William Petty (1623-1687) acuña el término “aritmética política” en su obra *Political Arithmetic*. En ella insta a la *medición* de los fenómenos sociales con iguales instrumentos que en las ciencias naturales. Además defiende la utilización gubernamental de los datos estadísticos que se obtengan.

John Graunt (1620-1674), un comerciante de tejidos que pasaría a ser considerado el primer demógrafo y fundador de la bioestadística. Introdujo las *tablas de esperanza de vida*, que expresan las probabilidades de supervivencia para cada grupo de edad. En su libro *Natural and Political Observations Made upon the Bills of Mortality*, publicado en 1662, analiza las tasas de mortalidad de Londres, en la Inglaterra de Carlos II, y la evolución de la peste bubónica en la ciudad, dando lugar a la primera estimación estadística de la población de Londres. Los “*bills of mortality*” (boletines de mortalidad) recogían las muertes registradas en las parroquias londinenses cada martes y, desde 1629, la causa de la defunción.

A Graunt se debe la primera *tabla de mortalidad* distribuida por edad, aunque no calculó la esperanza de vida que podría haber deducido de las tablas. Con posterioridad, estas *tablas* fueron desarrolladas por el astrónomo Edmund Halley (1656-1742) y sobre todo por el economista Gregory King (1648-1712), que contribuyó bastante al desarrollo de la *demografía* moderna. De manera especial su manuscrito *Natural and Political Observations and Conclusions upon the State and Condition of England*, de 1696, que describe las características demográficas de la población de Inglaterra y Gales (género, edad, estado civil, número de hijos...) e incluye proyecciones de la población.

- *Escuela estadística alemana*, también conocida como la *escuela estadística universitaria* porque se ubicaba en universidades alemanas (Berlín y Göttingen). Se diferencia de la escuela anterior por la promoción de la denominada “geografía política”. En concreto, la *comparación* de datos demográficos, socioeconómicos y políticos, relativos a diferentes países. Su promotor fue Ludwig von Seckendorff (1626-1692) y su fundador y representante principal Hermann Conring (1600-1681), quien introduce la *estadística* en un curso de ciencia política para describir datos sociopolíticos de los estados en estudios comparativos. Más tarde Gottfried Achewald (1719-1772) considera la estadística como disciplina independiente. La introduce como asignatura con el nombre de *estadística* (“*statistik*”) y la enfoca al análisis de datos del Estado. Su trabajo más conocido, *Staatsverfassung der Europäischen Reiche im Grundrisse*, de 1752, describe la agricultura, las manufacturas y el comercio, mediante estadísticas que permiten la comparación entre estados.

A estas aportaciones se suma la realizada por Adolphe Jacques Quételet (1796-1874). Un astrónomo y naturalista belga que se convirtió en uno de los propulsores más relevantes de la aplicación de los avances en la estadística al estudio de los fenómenos sociales. Colaboró en la creación de las *sociedades estadísticas* de Inglaterra y Francia, e impulsó la fundación de la *Asociación estadística internacional* (para la cooperación internacional en la recogida y el análisis de la información social). Pero, si por algo destaca Quételet, es por su *teoría de las regularidades de los fenómenos sociales*. Defiende que en la conducta humana, aparentemente azarosa, pueden observarse “regularidades” que pueden medirse mediante técnicas estadísticas de análisis. Razón por la cual se le considera “un pionero en la aplicación de modelos matemáticos en la sociología” (García Ferrando, 1979: 33).

Su obra *Sobre el hombre y el desarrollo de las facultades humanas: Ensayo sobre física social* (*L’homme et le développement de ses facultés, ou Essai de physique sociale*), publicada en 1835 y reeditada en 1869 (corregida, ampliada y con un título más breve: *Física social*), “marcó la transición de la simple descripción estadística al empleo consciente de los datos cuantitativos empíricos, para establecer las regularidades de la vida social” (Kovaliova, 1989: 136). En sus investigaciones socioantropológicas (*Sobre la ley del crecimiento del hombre e investigaciones sobre las tendencias criminales en distintas edades*), analiza las características físicas de las personas y las estadísticas de delincuencia. Para ello emplea el cálculo de las *probabilidades estadísticas*.

Quételet aplica la estadística a la criminología. Concibe el delito como un fenómeno social que puede conocerse y determinarse estadísticamente, porque se comete con cierta regularidad. Además desarrolla el concepto de “hombre medio”: el término medio de las características que resultan de la observación de diferentes sujetos.

A las aportaciones anteriores se suman otras pioneras en el ámbito de la *estadística* realizadas, entre otros autores, por:

- Pierre-Simon Laplace (1749-1827), astrónomo, físico y matemático francés, creyente del determinismo causal. Introduce el método de *mínimos cuadrados* (fundamental para la teoría de los errores) en su *Teoría analítica de las probabilidades* (*Théorie Analytique des Probabilités*), de 1820.
- Johann Carl Friedrich Gauss (1777-1855) profundizó en el *método de mínimos cuadrados*, en *ecuaciones diferenciales* y secciones cónicas. Entre sus publicaciones destaca *Disquisitiones Arithmeticae*, de 1801, con seis secciones dedicadas a la *Teoría de números*.
- George Udny Yule (1871-1951) realizó importantes contribuciones a la teoría y práctica de los análisis de correlación, regresión, covariación y series temporales. Entre sus publicaciones pioneras destaca *Introduction to the Theory of Statistics*, de 1911.

Las aportaciones desde la estadística para la fundamentación de la metodología cuantitativa confluyen con las efectuadas desde la investigación empírica. En concreto las primeras investigaciones sociales empíricas, de finales del siglo XVIII y principios del XX, de la mano de los *reformadores sociales*. Un conjunto de profesionales, de distintas disciplinas académicas (médicos, ingenieros, naturalistas), que querían aportar evidencia empírica de los problemas sociales existentes en su época. Su finalidad era eminentemente política: movilizar a la opinión pública para exigir medidas concretas de *reforma social*. Constituyeron *sociedades estadísticas*, que promovieron el desarrollo de las *encuestas sociales* en la investigación social. Entre las realizadas destacan:

- En Gran Bretaña: *El Informe estadístico de Escocia* (1791-1825), de John Sinclair; *Las condiciones morales y físicas de la vida de los obreros de la industria textil de Manchester* (1832), de James Kay-Schuttleworth, y *Vida y trabajo de los habitantes de Londres* (1889-1891), de Charles Booth.
- En Francia: *Higiene social* (1836) y *La prostitución en la ciudad de París* (1834), de A. Parent-Duchatelet; *Estadística sobre la moralidad en Francia* (1833), de G. Chanpneuf, y *Obreros europeos* (1877-1879), de Frederic Le Play. Además de los estudios del jurista André-Michel Guerry, *Ensayo sobre la estadística moral de Francia* (1832) y *La estadística moral en Inglaterra comparada con la de Francia* (1860), que contribuyeron al desarrollo de las estadísticas de delincuencia.

Éstos y otros estudios se describen en el capítulo 5, en el apartado correspondiente a los fundamentos teóricos y empíricos de la *encuesta* (a cuya lectura se remite).

Pero el divorcio entre la *teoría* y la *empiría* comienza a desvanecerse con la profesionalización del análisis de la realidad social desde ámbitos universitarios. Despuntan las aportaciones de los profesores alemanes Ferdinand Tönnies y Max Weber, y del sociólogo francés Emile Durkheim, el padre del *funcionalismo*.

1.1.2. La fundamentación teórica: del positivismo al funcionalismo

En 1989 Rodríguez Ibáñez sintetiza la evolución de la teoría sociológica en dos líneas fundamentales:

- *Estructural (dimensión Gesellschaft)*: va del *positivismo* clásico y el *evolucionismo* a la *teoría de sistemas*, pasando por Pareto y la reformulación sociológica de Durkheim y el *funcionalismo*.
- *Intersubjetiva (dimensión Gemeinschaft)*: arranca de movimientos no estrictamente sociológicos (como el *romanticismo* y la propuesta germana de *ciencias de la cultura*). Encuentra en Max Weber un refinador sociológico crucial, junto a otras figuras coetáneas como Tönnies y Simmel; años después, en los representantes de la *teoría crítica* y de las corrientes fenomenológicas.

La dimensión *estructural* predomina en los inicios de la sociología. Primero, bajo el nombre de *positivismo*, de la mano de Comte y de su maestro Saint-Simon. Después, con el nombre de *evolucionismo*, encabezado por Spencer. Culmina con el *funcionalismo* de Durkheim.

Auguste Comte (1798 -1857) fue, junto con el historiador Agustín Thierry, secretario del conde Henri de Saint-Simon durante siete años. De su colaboración deriva:

- Su teoría posterior de los tres estadios: *teológico* o ficticio (la infancia de la humanidad), el *metafísico* o abstracto y el *científico* o positivo, cuyo tránsito constituye el progreso de la humanidad. El conocimiento se orienta a la *experiencia* y la *observación*, a la deducción de sus *leyes generales* que permitan prever, controlar y dominar la naturaleza y la sociedad en provecho de la humanidad.
- La concepción de una nueva ciencia: la “física social”.

Comte toma de Montesquieu el determinismo de los fenómenos sociales e históricos, y de Condorcet la afirmación de que el progreso del espíritu humano constituye la razón del devenir de la sociedad. El objetivo principal de Comte era la reconstrucción del orden social, tras los cambios acaecidos con la Revolución francesa y la industrial, que demandaba el nacimiento de una nueva ciencia: una ciencia “positiva” de la sociedad. Tras la Revolución francesa, las concepciones metafísicas y teológicas del mundo habían perdido vigencia. A esa nueva ciencia comienza llamándola “física social”, por su equiparación a las ciencias de la naturaleza. Años después, en 1839, en el IV volumen de su obra *Curso de filosofía positiva (Cours de Philosophie Positivistic, 1830-1842)*, aparece con la que sería su denominación definitiva: “sociología”. Quería diferenciarla de la “física social” descriptiva de Quételet y hacerla una disciplina teórica, que tuviese como objeto el descubrimiento de las *leyes fundamentales* inherentes a los fenómenos sociales. Para ello tendría que basarse en dos postulados metodológicos fundamentales, en los que se asienta el *positivismo*:

- Los fenómenos *sociales* han de analizarse desde un punto de vista científico, a semejanza de los fenómenos *naturales*.
- Han de analizarse no como hechos aislados e independientes unos de otros, sino como fenómenos entrelazados y comprendidos en un todo.

En su *Curso de filosofía positiva* subraya muchos principios del *positivismo* aún vigentes en la actualidad. Destaca la *observación* como el *método* principal de la investigación social. A él añade el *experimento*, el *método comparativo* y el *histórico*. Este último consistirá en la comparación histórica de los distintos estadios consecutivos de la humanidad. Lamentablemente, sus reflexiones metodológicas no quedaron plasmadas en investigaciones empíricas concretas, aunque tuvo una gran repercusión en el desarrollo de la sociología.

Herbert Spencer (1820-1903), naturalista, filósofo y sociólogo británico, que introdujo el *darwinismo social* en Gran Bretaña y se convirtió en máximo exponente del *positivismo* y del *evolucionismo*. Su tesis fundamental es que todo en el mundo se rige por una ley de la evolución, tanto la materia orgánica como la propia sociedad. Su pensamiento se estructura en torno a dos ideas principales:

1. La comprensión de la sociedad como un *organismo* (asemeja los organismos sociales a los biológicos);
2. La *evolución social*.

Comparte con Comte:

- La *unidad de la ciencia*: la *sociología* es una ciencia *positiva*, basada en los mismos principios metodológicos de las ciencias naturales.
- El concepto de *ley científica*: todo fenómeno está sujeto a regularidades. Su identificación es el objetivo de cualquier análisis.
- El *holismo*: la comprensión de los hechos o fenómenos sociales como partes de un todo.
- La restricción del conocimiento a la *experiencia*, a lo observable.

A Spencer se debe el primer manual sobre el *concepto* y el *método* de la *sociología*, *The Study of Sociology*, de 1873; que después constituiría el prefacio de su voluminoso libro *The Principles of Sociology*, de 1877.

Vilfredo Pareto (1848-1923), economista, sociólogo e ingeniero italiano que, además de sus aportaciones en el área de la microeconomía, contribuyó en la fundamentación de la sociología con gran influencia en autores como Talcott Parsons. En su obra más conocida, *Trattato di Sociologia Generale*, de 1916, propugna la necesidad de reconstruir el *todo* social para alcanzar una explicación sintética del conjunto de la sociedad y de los movimientos que en ella se manifiestan. Nada de lo que sobrepasa el ámbito de la *experiencia* puede ocupar un lugar en la ciencia. La sociología ha de ser una ciencia empírica, inductiva y lógico-experimental, en la que debe imperar la neutralidad valorativa (al menos desde un punto de vista metodo-

lógico). Pero, pese a propugnar la necesidad de realizar investigación empírica, sus reflexiones, al igual que las de sus predecesores, no aparecen acompañadas de datos empíricos que las corroboren.

Ferdinand Tönnies (1855-1936), sociólogo alemán que en 1909 fundó la *Asociación alemana de sociología*. A diferencia de sus predecesores, sí realizó estudios empíricos, principalmente sobre conducta desviada y suicidio. Aunque no logra su conjugación con su teoría de la sociedad. Famosa fue su distinción entre *comunidad* y *sociedad* (*Gemeinschaft und Gesellschaft*), de 1887, que conforman diferentes tipos de relaciones sociales (la primera afectiva, de sentimientos y emociones; la segunda, racional, mercantil y mecanicista), según el tamaño de la población y su grado de complejidad en la división social del trabajo.

Émile Durkheim (1855-1917), sociólogo francés, que sí logra conjugar la *teoría* con la *empiría*. Con él culmina la visión *orgánica* de la sociedad y la consideración *positiva* de la sociología. Se convierte en el primer representante de la perspectiva *estructural-funcionalista*, y el primer autor que define el procedimiento que debe seguirse para alcanzar el conocimiento científico de los fenómenos sociales. Estableció formalmente la sociología como disciplina académica y es considerado, junto con Max Weber y Karl Marx, uno de sus padres fundadores.

En 1895 Durkheim crea el primer departamento de sociología en la Universidad de Bordeaux y en 1896 la primera revista dedicada a la sociología, *L'Année Sociologique* (en Estados Unidos, el primer curso que lleva el nombre de *sociología* lo dicta Sumner en la Universidad de Yale en 1876; la publicación de los dos primeros libros donde aparece la palabra *sociología* en su título data de 1854: *Sociology of South*, de Fitzbuth, y *Treatise on Sociology*, de Hughes. En 1893 Small ocupa la primera cátedra de sociología en la Universidad de Chicago; en 1895 aparece el primer número de la revista *American Journal of Sociology*; y en 1905 se constituye la *sociedad sociológica norteamericana*, presidida por Ward, fruto de su escisión de la *Sociedad económica norteamericana*. Después se sucede la creación de departamentos de sociología en diferentes universidades como las de Columbia, Wisconsin, Minnesota o Michigan, acentuando la especialización de la sociología, con un marcado carácter científico).

Durkheim perfeccionó el *positivismo* de Comte y promovió el método *hipotético-deductivo*. En 1895 publica *Las Reglas del Método Sociológico*, donde articula su concepción del *método* y del *objeto* de la *sociología*. Defiende la concepción de la sociedad como una suma de *hechos*, que han de analizarse como “*cosas*”. No porque los *hechos sociales* sean cosas materiales propiamente dichas, sino porque han de ser analizados como *data*, desde fuera, por su característica de observables y verificables empíricamente, a través del *método científico*. Lo cual exige:

- Desechar toda idea preconcebida. Sólo han de utilizarse criterios y conceptos científicamente contruidos. El sociólogo ha de aislar sus creencias político-religiosas para que no influyan en los *hechos sociales* que estudia.
- Buscar y recoger datos acordes al caso estudiado, y analizarlos de forma objetiva (alejándose de sensaciones de carácter subjetivo).

- Buscar primero su *causa*, definir el fenómeno social que le antecede, e indagar después en la *función* que cumple en el sistema social.

Para Durkheim el *método* idóneo es el *comparativo*, “el único que conviene a la sociología” (Durkheim, 1974: 133). Sólo podremos descubrir lo que es específico de una sociedad y lo que es común a la humanidad, si comparamos los mismos fenómenos en diferentes sociedades.

Sin duda uno de los mayores méritos de Durkheim ha sido el haber aplicado, de forma coherente, las *reglas del método* que propone a investigaciones concretas. Destaca *El suicidio*, de 1897, la primera investigación sociológica basada en datos y análisis estadísticos. También, *Las formas elementales de la vida religiosa*, de 1912, donde analiza la sociología de la religión y su relación con la teoría del conocimiento. De *El suicidio* se afirma que:

“Ha sido la primera obra importante que ha configurado la tradición de la teoría social con la aritmética política (...). Fue el prototipo perfecto de informe de investigación moderna (...). La parte teórica está bastante desarrollada, ya que el interés teórico del autor supera el problema del suicidio en sí. Las teorías anteriores son examinadas detalladamente, y el marco de la investigación está claramente definido; el significado de las estadísticas de suicidios es estudiado en todas sus ramificaciones; las hipótesis son modificadas y mejoradas como consecuencia de las confrontaciones con los datos, y reciben una nueva formulación al final de la obra” (Caplow, 1977: 10).

Entre los seguidores de Durkheim están los antropólogos funcionalistas Malinowski y Raddcliffe-Brown, que aplicaron las hipótesis durkheimianas (de la estructura normativa unitaria de la sociedad y su variación interna de funciones) a sus investigaciones sobre las comunidades primitivas. Al igual que sociólogos y psicólogos sociales como Parsons, Merton, Festinger, Duncan y un largo etcétera. Pero, no todos fueron seguidores. Sus tesis encontraron oponentes. Entre ellos el filósofo alemán Wilhelm Dilthey (1833-1911) y el sociólogo alemán Max Weber (1864-1920) y su sociología *comprensiva*.

A diferencia de Durkheim, Weber parte de la distinción propuesta por Dilthey en *Einleitung in die Geisteswissenschaften* (Introducción a la ciencia humana), de 1883, entre dos tipos diferentes de ciencia: *Naturwissenschaft* (ciencias naturales) y *Geisteswissenschaft* (ciencias del espíritu). Las primeras se centran en la observación de las manifestaciones externas de los fenómenos naturales. Las segundas, en la interpretación del *significado* y el *sentido* de la *acción* humana. Esta interpretación es posible gracias a un proceso de *comprensión* (*verstehen*): ponerse en el lugar del otro. Debemos comprender las razones personales o motivos que guían las decisiones a actuar de una determinada forma.

Weber se adhiere a la *escuela neokantiana*. Reconoce que se puede *comprender* una buena parte de las *acciones humanas* por *medios empáticos* (comprensión, intuición). Pero, prefiere considerar la *acción humana* como racional, orientada a *valo-*

res; e interpretar el sentido *subjetivo* inherente a la misma. La tesis de Rickert sobre los *conceptos* como medios para superar la diversidad de la realidad empírica se transforma, en la teoría de Weber, en la noción de *tipo ideal*: el “interés de la época”, expresado en forma de construcción teórica. No se extrae de la realidad empírica, sino que se construye como un esquema teórico “mediante el realce unilateral de uno o de varios puntos de vista y la reunión de una multitud de fenómenos singulares, difusos y discretos, que se presentan en mayor medida en unas partes que en otras o que aparecen de manera esporádica fenómenos que encajan en aquellos puntos de vista, escogidos unilateralmente, en un cuadro conceptual en sí unitario” (Weber, 1973: 79). La “construcción de tipos ideales abstractos no interesa como fin, sino exclusivamente como medio” (Weber, 1977: 63). Puede servir como patrón de referencia para comparar la realidad empírica. Si bien, puntualiza que el fin de la formación de conceptos de *tipo ideal* consiste no en tomar conciencia de lo genérico, sino de la particularidad de los fenómenos sociales. Mediante los *tipos ideales* se puede comprender la *motivación* de los actores y establecerse conexiones entre la *motivación* y los resultados de la *acción*.

Weber se centra en el aspecto *cualitativo* de los *hechos sociales*. Cuestiona la aplicación que del concepto de *ley* se ha hecho, tanto en las ciencias naturales como en las sociales. En su opinión, el establecimiento de *leyes* no constituye el objeto de la ciencia, sino que representa una fase previa al conocimiento que aspiramos. Cuatro serían los *pasos* a seguir en la adquisición del conocimiento:

1. Fijar *conceptos* y establecer *regularidades*.
2. Agrupar *causas* posibles en cada caso y el momento histórico determinado, señalando su importancia y a qué se debe.
3. Remontarnos al pasado para ofrecer una explicación *histórica* de las distintas particularidades de tales agrupaciones.
4. Evaluación de las posibles constelaciones de factores en el futuro (Weber, 1977: 41).

Tampoco acepta el principio de la *causalidad*: “el conocimiento de unas leyes de la causalidad no puede constituir el fin, sino el medio de estudio” (Weber, 1977: 45). Aunque sí comparte la defensa de la necesidad de que la sociología sea una *ciencia objetiva*, neutral y libre de valores. El científico se diferencia del político en que debe explicar *objetivamente* los hechos, sin implicaciones normativas o valorativas (en su obra *El político y el científico*, de 1919). Pero, a diferencia de Durkheim, no desarrolla su esquema metodológico general en una obra concreta. Su primer trabajo empírico data de 1892, y lo titula *La situación de los obreros de la industria textil en Alemania al este del Elba*. Se trata de una investigación empírica que Weber realiza en una fábrica textil alemana. Años más tarde, publica una introducción metodológica sobre aquella investigación (*Sobre el Problema de los Fundamentos Psicofísicos del Trabajo Industrial*). Sus intereses metodológicos coincidieron en dos temas fundamentales: “la elección de los informantes y la mejora del diseño del cuestionario” (García Ferrando, 1979: 62). En el capítulo 5 se detallan otras aportaciones de Weber en el ámbito de la

encuesta. Su obra fue determinante para el desarrollo de la *sociología cualitativa*. En concreto del *interaccionismo simbólico* posterior a Mead (Blumer, Glaser y Strauss), la investigación *fenomenológica* (de Schütz, Husserl, Berger y Luckman) y la *etnometodología* (o sociología cognoscitiva de Garfinkel, Douglas, Cicourel y Lindzey).

A caballo entre estas dos dimensiones del análisis de la realidad social se sitúa la obra de Karl Marx (1818-1883) y su método del *materialismo dialéctico*. Consiste en fijar las *leyes* que regulan los fenómenos sociales y, al mismo tiempo, enfatizar la perspectiva *histórica* (al igual que Weber). En su voluminosa e incompleta obra *El Capital*, Marx utiliza como fundamento de sus argumentaciones tanto datos estadísticos como documentos históricos. Su obra inspiró estudios clásicos como los de Lazarsfeld, Berelson *et al.* (1944) sobre la influencia de la clase media en el voto o los de Lockwood y Goldthorpe sobre el aburguesamiento de la clase obrera. Además de su influencia en sociólogos posteriores, como Luckàs o Gramsci (que desarrollaron aspectos inacabados de la obra marxiana), Althusser y Poulantzas (representantes de la versión estructuralista del marxismo), Horkheimer, Marcuse o Adorno (miembros de la *Teoría crítica* de la Escuela de Frankfurt), entre otros. Si bien, Paul F. Lazarsfeld (1901-1976) fue uno de los máximos exponentes del *neopositivismo* (que surgió en torno al “Círculo de Viena”) en la investigación social. En *Language of Social Research* (escrito junto con M. Rosenberg en 1955) defiende el estudio de los fenómenos sociales en términos de relaciones entre variables (dependientes e independientes), en la búsqueda de modelos causales. También que las teorías se pueden confirmar o invalidar de forma objetiva.

1.2. Características distintivas de la metodología cuantitativa versus cualitativa

Del apartado anterior habrá podido deducirse la existencia de dos “dimensiones”, *estructural* e *intersubjetiva* (Rodríguez Ibáñez, 1989), o “perspectivas” básicas, *cientificista/ cuantitativa* y *humanista/ cualitativa* (Alvira, 1983), de aproximarse al análisis de la realidad social. La primera fundamenta la metodología *cuantitativa*, mientras que la segunda, la *cualitativa*:

- a) La perspectiva “*cientificista/cuantitativa*” coincide con la dimensión *estructural* en evolución de la teoría sociológica (positivismo, evolucionismo, funcionalismo). Defiende la existencia de un único *método* común a todas las ciencias (el método de las ciencias naturales y exactas), la *medición* objetiva, la búsqueda de *leyes* generales, la demostración de la *causalidad* y la *generalización* de los resultados de la investigación.
- b) La perspectiva “*humanista/cualitativa*” corresponde a la dimensión *intersubjetiva* (romanticismo, Weber y la sociología “compreensiva”, la teoría crítica, la fenomenología). Defiende la *especificidad* de las ciencias sociales (rechaza el modelo metodológico de las ciencias naturales) y aboga por el análisis de lo individual y concreto, por medio de la *comprensión* o la interpretación de los

significados intersubjetivos de la acción social (desde el punto de vista del actor social). El énfasis se pone en el *lenguaje* y en los aspectos *micro* de la vida social (la interacción). Se centra en el individuo, en la descripción y comprensión de su conducta en su propio marco de referencia.

“El carácter reflexivo de la investigación cualitativa significa que los investigadores individuales inevitablemente inyectan algo de sí mismos en el proceso de investigación y, de este modo, en el resultado. En suma, las situaciones sociales nunca son suficientemente similares, a través del espacio y tiempo, para hacer la replicación posible” (Blaikie, 2010: 217).

A los métodos *cuantitativo* y *cualitativo*, Beltrán (1985) suma el *histórico*, el *comparativo* y el *crítico-racional*. La adecuación de cada uno dependerá de cuál sea el *objeto* de conocimiento en cada caso concreto.

Los métodos *comparativo* e *histórico* ya fueron destacados por el positivista Comte en 1839, en su *Curso de filosofía positiva*, como específico de la nueva ciencia, la sociología, junto con la *observación* y el *experimento*. Por su parte Durkheim, en *Las reglas del método sociológico*, de 1895, destaca el método *comparativo* como “el único que conviene a la sociología (Durkheim, 1974: 133). En cambio, el método *histórico* adquirirá un mayor relieve en la dimensión *intersubjetiva* del conocimiento sociológico. Y el método *crítico-racional* distingue a los autores englobados en la denominada *teoría crítica*, que defienden la consideración de los *valores* y los *finés sociales* como *objeto* de estudio en las ciencias sociales. A ellos se suman la *comprensión* de los *significados* y la *interpretación* o *hermenéutica*.

Para Ibáñez (1985) son tres las perspectivas en el análisis de la realidad social:

1. La *distributiva* (cuyo ejemplo más general y concreto es la encuesta social);
2. La *estructural* (el grupo de discusión);
3. La *dialéctica* (que aplica el componente *semiótico*, siendo el *socioanálisis* el ejemplo más general y concreto).

Esta tipología es modificada por Ortí (1989), quien engloba la perspectiva *estructural* y la *dialéctica* en una misma perspectiva, y separada de la *distributiva*. De esta forma, la clasificación de Ortí se asemeja a la anteriormente señalada como básica: la perspectiva *cuantitativa* y la *cualitativa*. El enfoque *distributivo* se corresponde con el método de las ciencias naturales, con la *explicación* y los métodos *cuantitativos*. Por el contrario, el enfoque *estructural-dialéctico* asume la especificidad de las ciencias sociales, la *comprensión* y los métodos *cualitativos*.

En suma, se distinguen dos “perspectivas” metodológicas fundamentales: la *cuantitativa* y la *cualitativa*. Ambas difieren en la estrategia seguida para la recogida de información y su análisis. Además proporcionan un marco filosófico y metodológico específico para el estudio de la realidad social. Por lo que, siguiendo a Filstead (1986), cabría su conceptualización como “paradigmas”. Precisamente Kuhn (1978) utilizó el término *paradigma* para referirse a este tipo de acuerdo entre *métodos* y *procedimientos*.

“Un *paradigma* representa “una *imagen* básica del objeto de una ciencia. Sirve para definir lo que debe estudiarse, las preguntas que es necesario responder, cómo deben preguntarse y qué reglas es preciso seguir para interpretar las respuestas obtenidas” (Ritzer, 1993: 598).

El cuadro 1.1 resume características distintivas que tradicionalmente se han atribuido a ambos *paradigmas*. El *cuantitativo* corresponde a la epistemología *positivista*, englobada en la dimensión *estructural* del análisis de la realidad social. En cambio, el paradigma *cualitativo* se asocia a la epistemología *interpretativa* (dimensión *intersubjetiva*), que se centra en el *sujeto* individual y en el descubrimiento del *significado*, los *motivos* y las *intenciones* de su *acción*. El concepto de “*verstehen*” de Weber se ajusta a esta última dimensión: el investigador debe procurar analizar el mundo social a través del sujeto que actúa para entender los *significados*, los *motivos* e *intenciones* de su *acción*.

CUADRO 1.1. Características diferenciales de los paradigmas *cuantitativo* y *cualitativo* en el análisis de la realidad social

	PARADIGMA CUANTITATIVO	PARADIGMA CUALITATIVO
<i>Base epistemológica</i>	Positivismo, evolucionismo, funcionalismo, teoría de sistemas, neopositivismo...	Romanticismo, historicismo, escuela neokantiana, fenomenología, interaccionismo simbólico, etnometodología...
<i>Concepción de la ciencia</i>	Unidad de la ciencia: analogía metodológica de las ciencias sociales con las ciencias naturales.	Diferenciación de las ciencias sociales de las naturales: singularidad de su objeto y método de conocimiento.
<i>Objeto</i>	La <i>cohesión estructural</i> y la <i>medición</i> precisa y objetiva.	La actividad <i>intersubjetiva</i> y procesos interactivos, dando relevancia a la condicionalidad histórica y su significado cultural.
<i>Énfasis</i>	Búsqueda de <i>regularidades</i> objetivas de los fenómenos sociales: identificación de <i>leyes universales</i> y <i>mediciones</i> válidas y fiables.	Análisis de lo <i>individual</i> y <i>concreto</i> . Énfasis en la variedad y especificidad de las sociedades humana y de sus manifestaciones culturales.
<i>Método</i>	Descripción y explicación de los <i>hechos sociales</i> como elementos integrantes de la sociedad global. Interesan los aspectos <i>cuantitativos</i> : <i>medición</i> objetiva, demostración de <i>causalidad</i> y <i>generalización</i> de los resultados de la investigación. Prevalencia del método <i>deductivo</i> (de la teoría a los datos).	<i>Comprensión</i> e <i>interpretación</i> de los agentes internos de la <i>acción social</i> . Destacan los aspectos <i>cualitativos</i> : descripción y comprensión de la conducta humana en su marco de referencia (individual y colectivo). Método <i>inductivo</i> (de los datos a la teoría).

CUADRO 1.1. (Continuación)

	PARADIGMA CUANTITATIVO	PARADIGMA CUALITATIVO
<i>Recogida de información</i>	<i>Estructurada</i> y sistemática. Máximo control y uniformidad en la recogida de datos y análisis para favorecer la comparabilidad (objetividad y replicabilidad).	<i>Flexible</i> : un proceso interactivo continuo marcado por el desarrollo de la investigación (dificulta la replicación)
<i>Análisis</i>	<i>Estadístico</i> para cuantificar la realidad social, las relaciones causales y su intensidad. Datos en forma de números. Análisis enfocados a la inferencia estadística (precisa del muestreo probabilístico).	<i>Interpretacional</i> , sociolingüístico y semiológico de los discursos, acciones y estructuras latentes. Datos en forma de palabras e imágenes. No precisa del muestreo probabilístico, sino que los sujetos aporten información relevante.
<i>Alcance de los resultados</i>	<i>Nomotética</i> : búsqueda cuantitativa de leyes generales de la conducta.	<i>Ideográfica</i> : búsqueda cualitativa de significados de la acción humana.

Pero, como suele ser habitual cuando se trata de resumir características distintivas, puede errarse de excesiva simplificación. Para Brannen (1994: 6), “muchas de la investigación cualitativa es simplemente descriptiva”. Igualmente, la investigación *cuantitativa* no siempre se enfoca a la comprobación de *hipótesis* (deducidas de *teorías* existentes). Con frecuencia se dirige sólo a la *descripción* de hechos sociales, opiniones o actitudes individuales, sin ninguna intención explicativa. Por su parte, aunque sea cierto que la investigación *cualitativa* busque entender el punto de vista de la persona analizada, es –en palabras de Hammersley (1994: 45)– “muy raro que la investigación cualitativa se restrinja a documentar el innato punto de vista”. Por lo cual, hay que remarcar el carácter genérico de las características que resume el cuadro 1.1. Asimismo ha de subrayarse que, “aunque existen metodólogos que se adhieren a los extremos de o la filosofía *positiva* o a la *interpretativa*, existen también muchos que caen en el medio de este *continuum* metodológico y filosófico” (O’Connell y Layder, 1994: 32).

“Los estudios cuantitativos pueden recoger algunos datos en palabras (por ejemplo, preguntas abiertas en una encuesta, o texto en el que se emprende un análisis de contenido). Por medio de algún proceso de codificación, estos datos se transforman en forma numérica. De manera similar, algunos estudios cualitativos pueden producir simples tablas de frecuencias y porcentajes para resumir algunas de las características de los datos no numéricos” (Blaikie, 2010: 215).

El enfrentamiento entre ambos paradigmas marcará el desarrollo de la teoría y la investigación social a lo largo del siglo XX. Primero bajo el predominio de la in-

vestigación *cualitativa*, bajo el paraguas de la *Escuela de Chicago* y estudios de gran alcance como *El campesino polaco en Europa y América* (1918-1920), de W. Thomas y F. Znaniecki. En él se analiza la adaptación social de los inmigrantes polacos en Estados Unidos a principios del siglo XX a través de diferentes fuentes de datos: 754 cartas de inmigrantes polacos en Estados Unidos; archivos de un periódico adquirido por Thomas en Polonia, y que cubría un período de 20 años; archivos de parroquias polacas en Estados Unidos; documentos de algunas organizaciones de ayuda a inmigrantes polacos; y la autobiografía de un joven inmigrante polaco llamado Wladeck; todo lo cual convierte al estudio en original y pionero en la aplicación del *método biográfico*.

La revista *The American Journal of Sociology* se convierte en el órgano de divulgación de los estudios de la Universidad de Chicago, entre los que sobresalen: *The Gang*, de Tharsher; *The Jack Roller* y *Brothers in Crime*, de Shaw; y *The Ghetto*, de Louis Wirth. Estudios que comparten una misma metodología. A la *observación participante* (considerada la técnica principal) añaden otras fuentes: *informes oficiales* (municipales, de los tribunales de justicia, de la policía, de las asociaciones locales), *datos censales*, *documentos privados* y *biografías*, *artículos de prensa*, *informes de especialistas*, *entrevistas* y *encuestas*, fundamentalmente. Se suman otros estudios de gran alcance, como los realizados por el matrimonio Lynd: *Middletown* (1929 y 1937); la serie *Yankee City* (1941-1959) de W. Lloyd Warner, o *Street Corner Society* (1943) de William Foot Whyte, que significó la culminación de la técnica de *observación participante* en la investigación social.

Con la llegada a EE UU, a finales de los años treinta y principios de los cuarenta, de algunos miembros del *Círculo de Viena* (huyendo de la persecución nazi), se produce un cambio de rumbo. La confluencia del *positivismo lógico* del *Círculo de Viena* con el *pragmatismo* americano, junto con las críticas manifestadas contra la metodología *cualitativa*, supusieron el fin del dominio de la *Escuela de Chicago* y el auge de la metodología *cuantitativa*. La Universidad de Columbia tomó el relevo a la de Chicago, al hallarse en ella los dos sociólogos principales del momento: Robert K. Merton y Paul F. Lazarsfeld. Además de aportaciones anteriores como las realizadas en el campo de la medición de las actitudes (caso del artículo de L. L. Thurstone “*Attitudes can be measured*”, de 1928, y otros referenciados en el capítulo 2).

A principios de los sesenta, el *funcionalismo* pierde la hegemonía entre los jóvenes sociólogos. Surgen nuevas perspectivas de la realidad social y renacen viejas teorías (algunas de ellas con nombres nuevos). Si bien, ninguna perspectiva llega a dominar el desarrollo de la sociología del momento. Las antiguas disputas entre *positivistas* e *historicistas* resurgen de nuevo. Sirva de ilustración lo manifestado por Adorno al respecto:

“Hace mucho tiempo que la sociología empírica se ha dado cuenta de lo que pierde en contenido específico por culpa de la generalización estadística. A menudo viene contenido en un detalle algo decisivo desde el punto de vista de lo general y que se escapa a la mera generalización. De ahí la fundamental necesidad de

contemplar los sondeos estadísticos por medio de *case studies*. El objetivo de los métodos sociológicos cuantitativos debería ser, asimismo, la comprensión cualitativa; la cuantificación no es un fin en sí misma, sino un medio para dicho fin” (Adorno *et al.*, 1973: 51-52).

Adorno subordina la *cuantificación* a la *comprensión cualitativa* y se encuadra, junto con Horkheimer, Marcuse, Bloch, o Habermas, dentro de la *teoría crítica* (donde converge el pensamiento de Marx, con el de Weber y Freud). Se caracteriza por la crítica a la exaltación de la *experiencia* y de las ciencias empírico-analíticas, y la defensa de la *autorreflexividad*. Sus teóricos defienden la necesidad de que la ciencia social trascienda la *aparencia* de los fenómenos sociales, para ir hasta la *esencia* de las cosas. Los orígenes de esta corriente crítica se remontan, no obstante, a los años treinta, cuando se constituye el *Instituto de investigación social de Frankfurt*, en torno a las figuras de Horkheimer (1893-1973), Adorno (1903-1969) y Marcuse (1897-1968), principalmente. De 1937 data la obra de Horkheimer *Teoría tradicional y teoría crítica*, un manifiesto de la *teoría crítica* donde el autor define qué debe entenderse por *teoría crítica*.

En los años sesenta Jürgen Habermas recoge el legado de los representantes de la *Escuela de Frankfurt*, pero en el marco de la crítica al *estructural funcionalismo*. En *Theory and Practice* (1974), Habermas sostiene que el *paradigma* ya no es la *observación cuantitativa*, sino el *diálogo*. En disciplinas *hermenéuticas*, como la sociología, ha de recurrirse al *diálogo* y no a los procedimientos de *medición*.

En los años sesenta también se recupera la perspectiva del *interaccionismo simbólico* de la *escuela de Chicago* (Mead, Cooley, Thomas, Park), de la mano de algunos de sus antiguos miembros. Entre ellos destacan Blumer, Becker, Hughes, Harnowitz y Goffman, para quienes la sociedad es *interacción*.

“El estudio de la interacción habrá de hacerse desde el punto de vista del actor. Dado que la acción la construye el actor a partir de lo que percibe, interpreta o juzga...” (Blumer, 1966: 542).

Una reseña de las investigaciones realizadas desde el *interaccionismo simbólico* ha de incluir las aportaciones de Turner (1962) a los problemas de identidad, el proceso de *role-making* y los movimientos sociales colectivos; los estudios de Rose (1962) sobre el poder; o las investigaciones sobre el proceso de *estigmatización* y el desarrollo de la carrera delictiva de los llamados teóricos del *etiquetamiento*: Becker (1963), Lemert (1969), Erikson (1962), Kitsuse (1962) o Schur (1968), entre otros. Otros sociólogos se orientarán hacia versiones de la *fenomenología* (impulsada en obras como *Ideas para una fenomenología pura* y *Filosofía fenomenológica*, de Edmund Husserl, de 1913). Caso de Peter L. Berger y Thomas Luckman y *La construcción social de la realidad* (1968), que introduce la noción de “construcción social”: la realidad se construye socialmente y la sociología del conocimiento ha de analizar los procesos por los cuales esto se produce.

En los años sesenta también surge, en la Universidad de Los Ángeles, una nueva corriente de pensamiento denominada *etnometodología*, de la mano de Harold Garfinkel (y su *Studies in Ethnometodology*, de 1967). El estudio del *lenguaje* (descifrar los significados culturales) adquiere una gran importancia. Los *etnometodólogos* analizan cómo las personas definen y construyen, cara a cara (en interacción), cada situación social. No rechazan el *empirismo*. Si bien proponen una nueva técnica: la *simulación*. Consiste en partir de una situación que se cree normal y, desde ella, intentar crear confusión. Por lo que su operatividad en contextos que no sean microsociológicos se estima dudosa. La *etnometodología* fue fructífera en ámbitos concretos como la *sociolingüística* (análisis del *discurso*) y la *psicología social* de instituciones (juzgados, hospitales, escuelas, comisarías), en la ilustración de la realidad cotidiana de su funcionamiento.

Pero éstos no son los únicos desarrollos teóricos de los años sesenta. Junto a ellos conviven otros, como la *teoría de sistemas* (una actualización del *paradigma normativo*) o la *teoría del intercambio* (aunque los orígenes de esta última se remontan a 1907, cuando Georg Simmel publica un artículo que lleva por título el término “intercambio”).

En suma, la década de los años sesenta supuso el final de la hegemonía del paradigma *cuantitativo*, o *positivista*, y el renacer del paradigma *interpretativista* o *cualitativo*. A ello contribuyó la reaparición de las ideas fenomenológicas, junto con la nueva visión de la ciencia y del desarrollo del conocimiento científico propiciado por alumnos de Karl R. Popper (1902-1994): Imre Lakatos (1922-1974) y Thomas S. Kuhn (1922-1996). Ambos contribuyeron a la reafirmación del pluralismo teórico-metodológico en las ciencias sociales.

En este resurgir fue clave la fecha de 1967 (Tesch, 1992). En ese año se publicaron obras de gran repercusión en el desarrollo posterior de la sociología, como *Studies in Ethnometodology*, de Garfinkel, o *The Discovery of Grounded Theory*, de Barney Glaser y Anselm Strauss. Un año antes (en 1966) había aparecido la obra de Severyn T. Bruyn, *Human Perspective in Sociology*, y la de Webb, Campbell, Schwartz y Sechrest, *Unobtrusive Measures: Nonreactive Research*. Dos años después (en 1969) se edita el libro de Herbert Blumer, *Symbolic Interactionism*. Si bien, este último autor ya había publicado (también en 1967) un artículo sobre el mismo tema. A partir de entonces, se multiplican las publicaciones sobre metodología *cualitativa*. Con ello se trata de llenar los huecos descubiertos en el paradigma *cuantitativo*. Éste se muestra ahora insuficiente para cubrir todas las demandas explicativas que exige el análisis de la realidad social.

A partir de los años setenta, comienza a hablarse de *crisis* de la sociología. La inexistencia de un *paradigma* dominante, con el consiguiente enfrentamiento de los enfoques teórico-metodológicos existentes (en busca de hegemonía o supremacía sobre los demás), provoca la *crisis* y la consiguiente redefinición de su objeto de conocimiento. Crisis que, favorablemente, trae un período de pluralidad paradigmática en la sociología. Se reconoce la pluralidad de vías para acceder a la realidad social. No se trata de afirmar un *paradigma* sobre otro, sino de buscar compatibilidades

entre ellos. La frase de Reichardt y Cook (1979: 27), “Es tiempo de dejar de construir muros entre métodos y empezar a construir puentes”, ejemplifica esta predisposición a la *complementariedad* entre *paradigmas*, y no a su oposición. Como consecuencia, un nuevo concepto se pone de moda, sobre todo en los años ochenta: la *triangulación* o utilización de distintos métodos en una misma investigación. Un concepto que tiene sus raíces en ámbitos de conocimiento diferentes a la investigación social (la topografía, la navegación e, incluso, en la logística castrense), en los que adquiere un mismo significado: la utilización de múltiples puntos de referencia para localizar la posición exacta de un objeto en el espacio. De esta forma se logra una mayor precisión que la alcanzada mediante la aplicación de un único punto de referencia. En las ciencias sociales, el término *triangulación* adquiere un significado similar. Por él se entiende la aplicación de distintas metodologías en el análisis de una misma realidad social. El concepto fue introducido en la obra de Webb *et al.* (1966), antes mencionada, aunque adquiere un mayor desarrollo en 1970, tras la publicación de *The Research Act* de Norman K. Denzin.

1.3. La articulación metodológica

La conjunción de diferentes métodos de recogida de información en una misma investigación ya estaba presente en los mismos orígenes de la investigación social. Sea el caso del pionero y famoso estudio de Charles Booth *The Life and Labour of the People of London* (1889-1903). Una voluminosa investigación (cuya primera edición se publica en dos volúmenes en 1889, la segunda en nueve volúmenes de 1892 a 1897 y la tercera en diecisiete volúmenes de 1902 a 1903), donde se describe, con minuciosidad, las condiciones de vida de la población trabajadora londinense. La información se extrae de diversas fuentes: los *censos* de población realizados hasta la fecha en Inglaterra (el primero data de 1801), de *informes* (policiales, de organizaciones de beneficencia, de inspectores escolares y sanitarios), de la *observación* directa de barrios obreros y la realización de *entrevistas abiertas* a hombres de negocios y asistentes sociales. A partir de lo cual se elaboraron *mapas* que describían el grado de pobreza de diferentes barrios londinenses. Pero también de los realizados en la *Escuela de Chicago*, que combinaban el *estudio de casos* con métodos *históricos* y *estadísticos*.

Pero la necesidad de articular distintos métodos en una misma investigación no se plantea a fondo hasta los años cincuenta, cuando se debate la conveniencia de combinar la *encuesta* con el *trabajo de campo* cualitativo (Vidich y Shapiro, 1955; Zelditch, 1962; McCall y Simmons, 1969; Sieber, 1973); o técnicas concretas de recogida de información como la *entrevista* y la *observación participante* (Becker y Ger, 1958; Trow, 1957).

A finales de los cincuenta, en 1959, Campbell y Fiske escriben un artículo, que iba a tener una gran repercusión, en la revista *Psychological Bulletin* (“Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix”). En él concep-

túan la *operacionalización múltiple*: la necesidad de utilizar más de un método en el proceso de medición para, de esta forma, aumentar la *validez* de los hallazgos y el grado de confianza en los mismos. Por *validez* entienden el grado de adecuación de una medida particular respecto a la realidad a la que apunta. En su opinión, la adecuación de cualquier *constructo* hipotético debe establecerse por su *validez convergente* (el acuerdo entre los distintos métodos de medir el mismo constructo) y la *discriminante* (la diferenciación entre los resultados cuando se emplea el mismo método para asegurar cualidades hipotéticamente diferentes de los mismos objetos).

En 1966, Webb y otros autores (entre los cuales figura Campbell) publican *Unobtrusive Measures: Nonreactive Research in the Social Sciences*, donde por vez primera aparece el término “triangulación”. Basándose en las ideas de Campbell y Fiske (1959), estos autores (Webb *et al.*, 1966: 3) defienden la necesidad de aplicar una “triangulación de los procesos de medición”. Con ello se persigue el logro de un mayor grado de *validez* de los hallazgos de una investigación. En 1970, Denzin dedica un apartado de su obra *The Research Act* (la Parte IV) a la definición de la *triangulación* en la investigación social. Diferencia la *triangulación* de datos, investigadores, teorías y metodologías:

- *Triangulación de datos*: la modalidad más habitual en la práctica de la investigación social. Consiste en la utilización de varias y variadas fuentes de información sobre un mismo objeto de conocimiento, con el propósito de contrastar la información recabada. Cuando los datos convergen, la evidencia adquiere mayor fuerza.

Por ejemplo, en mi tesis doctoral (Cea D’Ancona, 1992), la medición de la *reincidencia* en conducta delictiva se hace conjugando tres fuentes de información: el expediente del menor (detenciones policiales y de la Guardia Civil, *denuncias* efectuadas por familiares u otras personas), la información proporcionada por la Dirección General de la Policía (detenciones policiales y de la Guardia Civil posteriores a la intervención del Tribunal de Menores) y la facilitada por la Dirección General de Instituciones Penitenciarias (reclusión en centros penitenciarios y situación procesal).

- *Triangulación de investigadores*: el equivalente a lo que se conoce por *equipos interdisciplinarios*. Están integrados por investigadores de distintas áreas de conocimiento o especialistas en diferentes *metodologías*, que analizan un mismo objeto de estudio desde ángulos distintos.
- *Triangulación teórica*: analizar un mismo objeto desde perspectivas teóricas diferentes. Ello implica tomar una unidad común para la observación del problema de investigación, extraer diversas proposiciones específicas de cada perspectiva teórica, operacionalizar los conceptos comprendidos en las proposiciones, recoger la información y contrastar cada proposición con los datos recabados.

Por ejemplo, en el estudio de la delincuencia juvenil, conjugar las perspectivas teóricas que enfatizan el nivel *macroestructural* (como la *Teoría de la subcultura* de

Albert K. Cohen o la *Teoría de la oportunidad diferencial* de R. A. Cloward y L. E. Ohlin) con las que indagan en los procesos *microinterpersonales* (como la *Teoría del etiquetamiento* de Howard Becker y Edwin M. Lemert). Su articulación favorecería una visión más completa de las causas de la delincuencia.

- *Triangulación metodológica*: la última forma genérica de triangulación. Comprende la triangulación *intramétodo* (el investigador escoge un único método o estrategia de investigación, pero aplica distintas técnicas de recogida y de análisis de datos) y *entre métodos* (combinación de métodos no similares para analizar un mismo objeto de estudio y paliar las limitaciones de cada uno). La última es la modalidad “más satisfactoria” (Denzin, 1975) y “más popular” (Jick, 1979), convirtiéndose en el vehículo para la *validación cruzada* (cuando se alcanzan los mismos resultados con métodos distintos, pero que apuntan a la misma dimensión del problema de investigación).

Denzin (1970/1975) defiende la *triangulación multimétodo* como ideal. Consiste en la articulación, en un mismo estudio, de los cuatro tipos de triangulación. Lo que permitiría obtener resultados más satisfactorios que si, únicamente, se hubiera triangulado por fuentes de *datos*, *investigadores*, *teorías* o *métodos*. A partir de entonces se suceden las publicaciones que tratan la *triangulación*. Entre ellas, el artículo de Jick, “Mixing Qualitative and Quantitative Methods: Triangulation in Action”, de 1979; o, el magnífico libro de Brewer y Hunter, *Multimethods Research: a Synthesis of Styles*, de 1989, que especifica actuaciones para el ejercicio, con éxito, de la *triangulación*:

- a) Una *triangulación* exitosa requiere un análisis cuidadoso de cada *método*, en relación con los otros *métodos*, y respecto a las demandas del proceso de investigación.
- b) Qué *métodos* y cuántos *métodos* han de emplearse depende de la información que se precise para esclarecer el problema de investigación.
- c) Muchos estudios *multimétodo* tienen objetivos mezclados. Se puede generar una teoría con un *método* y comprobarla con otro.
- d) Tanto al crear como al comprobar una teoría, ciertas variables pueden ser accesibles a un *método*, pero no a otros. Por lo que se precisa una serie de *métodos* de recogida de datos, bastante vinculados, para considerar la significatividad de diferentes tipos de variables.
- e) Hay que procurar que los distintos *métodos* que se apliquen para obtener información midan realmente los mismos *conceptos teóricos*. Ello es necesario, si se pretende que los resultados obtenidos (mediante diferentes procedimientos) sean comparables.
- f) Para que la aplicación de un *método* no influya en los resultados alcanzados por otro *método*, sería conveniente seguir alguna de las opciones siguientes:
 - Realizar la *recogida de datos escalonadamente*, de manera que la información que sea más vulnerable a la influencia del investigador o de los suje-

tos analizados (generalmente aquella obtenida mediante *métodos reactivos*, como la *encuesta*) sea recabada primero; y después las menos reactivas (como el uso de *fuentes documentales y estadísticas*).

Por ejemplo, para la evaluación de la actuación de un centro, aplicar primero la *observación*, después la *encuesta* y, por último, acceder a los *informes* registrados en el centro. De esta forma se evitaría el efecto que el conocimiento del acceso (por parte del equipo investigador) a los *informes* del centro pudiese afectar a la respuesta de la *encuesta*.

- *Dividir la muestra en submuestras*, o *extraer varias muestras*, y en cada una de ellas, aplicar una técnica distinta de recogida de información. Ello evitaría, igualmente, el efecto directo de un *método* sobre otro, al no exponerse a los mismos sujetos a más de un *método*. Además, podrían comprobarse los efectos concretos de un *método* en otro. Para lo cual se compararían los resultados obtenidos mediante este procedimiento con aquellos alcanzados después de exponer a algunos sujetos (extraídos de las *muestras* de forma *aleatoria*) a ambos tratamientos.

Pero, téngase presente que este procedimiento de extracción de varias *muestras* no garantiza la inmunidad frente a los *sesgos* derivados del conocimiento, por parte de los integrantes de una de las muestras, de las experiencias de investigación de las otras *muestras*. De ahí que, si se desea evitar este tipo de *sesgo*, sea necesario bien aplicar cada una de las técnicas de recogida de datos escalonadamente, o bien proceder de forma rápida y simultánea en las distintas *muestras*.

Los argumentos de Denzin (1970) a favor de la *triangulación* (“El sociólogo debe aprender a emplear múltiples métodos en el análisis del mismo evento empírico”, p. 13) alcanzan una gran repercusión en diferentes campos de conocimiento, como la *evaluación de programas* (Reichardt y Cook, 1979; Schwartz y Jacobs, 1979; Patton, 1980; Madey, 1982) o en la *etnografía* (Hammersley y Atkinson, 1983; Burgess, 1984; Hammersley, 1992, entre otros). Se defiende la existencia de diseños de “métodos múltiples” o “métodos mezclados” (Teddle y Tashakkori, 2003; Fielding, 2010; Hesse-Biber, 2010).

“Una investigación de métodos mezclados es un diseño de investigación con supuestos filosóficos al igual que métodos de indagación. Como metodología, implica supuestos filosóficos que guían la dirección de la recogida y el análisis de los datos y la mezcla de aproximaciones cuantitativas y cualitativas en muchas fases del proceso de investigación. Como método, se centra en la recogida, el análisis, y la mezcla de datos tanto cuantitativos como cualitativos en un único estudio o series de estudios. Su premisa fundamental es que el uso de aproximaciones cuantitativas y cualitativas en combinación proporciona una mejor comprensión de los problemas de investigación que cualquier aproximación sola” (Cresswell y Plano Clark, 2007: 5).

En el otoño de 2005, surge una nueva revista, *Journal of Mixed Methods Research* (publicada por Sage), que se dedica exclusivamente a la publicación de estudios de métodos mezclados. Y, como destaca Hesse-Biber (2010), los avances tecnológicos están contribuyendo a aumentar el interés en combinar nuevas prácticas analíticas y diseños de métodos mezclados. Programas de software de análisis de datos cualitativos (CAQDAS), *por ejemplo*, pueden incorporar datos cuantitativos en sus programas (por lo que pueden crearse medidas cuantitativas de datos cualitativos). Si bien, también aparecen detractores (Silverman, 1985; Fielding y Fielding, 1986; Blaikie, 1991). Estos últimos reconocen que la *triangulación* favorece la adquisición de un conocimiento más amplio y profundo de la realidad social, pero niegan que este conocimiento sea necesariamente más objetivo y fiable que el obtenido mediante una única *metodología*. Las ventajas e inconvenientes de la *articulación* metodológica se resumen en el cuadro 1.2.

CUADRO 1.2. Ventajas e inconvenientes de la articulación metodológica

VENTAJAS	INCONVENIENTES
Obtención de una información más profunda y diversa, al cubrirse más dimensiones de la realidad social.	Conlleva un mayor coste económico y temporal.
Incita al uso de múltiples puntos de vista o paradigmas.	Dificultad de encontrar una unidad de observación común a la que aplicar métodos diferentes e investigadores con formación en ellos.
Mayor <i>validez</i> de los resultados de la investigación (cuando se llega a los mismos hallazgos con métodos diferentes).	Mayor dificultad para repetir la investigación.
La <i>teoría</i> queda más reforzada, en caso de ser confirmada. Si no lo es, proporciona un mayor fundamento para su modificación.	La aplicación de una técnica de recogida de información puede <i>sesgar</i> los datos que recaben técnicas posteriores.
Enriquece el diseño y los resultados de la investigación.	Problemas de <i>comparabilidad</i> de los resultados, al verse afectados por diferentes fuentes de error.
	Existencia de incompatibilidad epistemológica entre algunos métodos.

Mediante la articulación metodológica tratan de solventarse las limitaciones de cada método con las potencialidades del otro.

Por ejemplo:

- La *encuesta* aparece como el método o estrategia de investigación más adecuada, cuando se precisa información (hechos, opiniones, actitudes) de un

gran número de personas, localizadas en un área geográfica amplia, en un breve período de tiempo, y a un precio relativamente bajo.

- El *estudio de casos*, cuando se desea una información más profunda de la realidad social. En especial, de las relaciones sociales.
- La *experimentación* se centra en la comprobación de *hipótesis causales*.
- El *uso de datos disponibles* se muestra más adecuado para indagar en hechos del pasado y procesos de cambio social. Si bien, en ambos casos se aconseja combinar la información proporcionada por esta *estrategia* con otras *estrategias de investigación*.

Tradicionalmente, el paradigma *cuantitativo* ha relegado al *cualitativo* al nivel que Reichenbach (1973) denomina “contexto de descubrimiento”. Se caracteriza por la realización de actividades *exploratorias*, que contribuyen a la generación de *hipótesis*. Por el contrario, los *métodos cuantitativos* se reservan el “contexto de la justificación” científica de los hallazgos, su significatividad. O, como señala Blaikie (2010: 215),

“Los métodos cualitativos pueden ser usados por investigadores cuantitativos en un estadio exploratorio para sugerir hipótesis o facilitar el desarrollo de instrumentos de investigación. Los métodos cualitativos son vistos como suplementarios a los métodos cuantitativos”.

Desde su experiencia en la investigación social comparativa, Allardt (1990) argumenta que por medio de datos *cuantitativos* el investigador es capaz de explorar los fenómenos estudiados. Si bien, para que los resultados obtenidos sean más interpretables, es preciso que se efectúen estudios *cualitativos*, mediante los cuales pueda profundizarse en el tema de estudio. De igual manera se pronuncian otros autores. Para Walker (1988: 22), “la investigación cualitativa, más generalmente, puede ayudar a interpretar, iluminar, ilustrar y cualificar las relaciones determinadas estadísticamente”.

No obstante, la práctica de la investigación social demuestra que lo inverso también es cierto. Ambas metodologías se adecuan a cualquier fase del proceso de investigación social. De lo que se trata no es de subordinar unas a las otras, sino de *articularlas* o combinarlas en un mismo proyecto de investigación, en función de cuál sea el objetivo concreto a cubrir. Como subraya Beltrán (1989: 40),

“Basta con afirmar el método cualitativo junto al cuantitativo, dejando que sea el propio objeto de conocimiento el que lo justifique y reclame en función de sus propias necesidades, perfectamente diferenciadas”.

La *complementariedad paradigmática* se impone como “una necesidad práctica impuesta por la investigación” (Pérez Agote, 1989: 118). De la articulación de la *en-cuesta* con otras estrategias de investigación se habla en el capítulo 5.

EJEMPLO DE ARTICULACIÓN METODOLÓGICA

Mikolaj Stanek (2011: “Nichos étnicos y movilidad socio-ocupacional. El caso del colectivo polaco en Madrid”, *Reis* 135: 69-88).

El estudio analiza la relación entre la participación en los nichos étnicos y la movilidad socio-ocupacional de los inmigrantes polacos en la Comunidad de Madrid. Para ello se combina el análisis estadístico de datos de *encuesta* con el análisis de los discursos recogidos mediante *entrevistas en profundidad*. Los datos estadísticos proceden de la “Encuesta de inmigrantes polacos en la Comunidad Autónoma de Madrid” a una muestra de 210 personas de origen polaco mayores de 17 años, residentes en la Comunidad Autónoma de Madrid, y realizada por el autor en el período de abril a septiembre de 2006. La *encuesta* recoge información sobre diversos aspectos de la vida laboral (tipo de contrato, ocupación, sector de actividad, duración de cada empleo) en diversos momentos de la trayectoria migratoria (el momento de partida, el primer empleo en España y el empleo actual). La aplicación del análisis de *regresión logística* permite identificar los factores que influyen en la mejora de posiciones en la estructura socio-ocupacional.

El estudio se completa con el análisis de los resultados de las *entrevistas en profundidad* a 26 inmigrantes polacos. Las entrevistas tenían carácter semiestructurado. Los temas de conversación giraban en torno a las trayectorias migratorias y laborales de los inmigrantes. Se intentó que reflejasen la diversidad de situaciones en las que se encuentran los inmigrantes, comprender sus actitudes y actuaciones, así como reconstruir los cambiantes contextos sociales y económicos en los que llevan a cabo sus estrategias. El objetivo de esta parte del estudio era facilitar material interpretativo para la explicación y comprensión de las regularidades en las relaciones causales observadas en el análisis cuantitativo previo. Se quería complementar el marco explicativo situando las regularidades observadas mediante la inferencia estadística en un amplio contexto, configurado tanto por los condicionamientos estructurales como por las estrategias y motivaciones individuales.

Los resultados ponen de manifiesto que los polacos se concentran en dos nichos étnicos específicos: los varones en la construcción y las mujeres en el servicio doméstico en régimen por horas. En este contexto, la movilidad socio-ocupacional está fuertemente diferenciada en razón de género, dependiendo de la distribución de los papeles productivos y reproductivos entre varones y mujeres dentro de los hogares y de las estructuras de oportunidad existentes dentro de cada nicho étnico.

1.4. Del método científico al proyecto de investigación

La metodología cuantitativa se asienta en el *método científico*, primeramente configurado por Francis Bacon (1561-1626) y su defensa del *empirismo* como característica distintiva de la ciencia. Desestima la *lógica* como fuente de verdad (*método deductivo*), anteponiendo la *experiencia*, la *observación*, como paso previo a la generalización y la formación de teorías (*método inductivo*). Los científicos han de

ser escépticos y no aceptar explicaciones que no se puedan probar mediante la observación y la experiencia (*empirismo*).

En 1934, Karl Popper rechaza la concepción de la ciencia como saber basado en la *acumulación inductiva*, en su célebre obra *La lógica de la investigación científica* (*Logik der Forschung*). “La base empírica de la ciencia objetiva no tiene nada de absoluta” (Popper, 1977: 106). Lo que confiere carácter de *cientificidad* es la solidez o fortaleza *hipotético-deductiva* de la *teoría*, su capacidad de ser rebatida por la *experiencia*. La clave del método científico se sitúa en el *falsacionismo*. Las teorías se convierten en “conjeturas audaces que serán abandonadas tan sólo si entran en conflicto con posteriores observaciones” (Popper, 1967: 58).

Los planteamientos de Popper fueron profundizados por sus discípulos. Entre ellos, Imre Lakatos y su propuesta de *falsacionismo refinado* o *sofisticado*: para cada *teoría científica* se trata de “especificar los hechos que la confirman y la probabilidad de la teoría a la luz de estos hechos” (Lakatos, 1975: 208). Su aceptación dependerá de si posee (respecto a otras *teorías* rivales o que la precedan) un contenido empírico excedente probado, que facilite el descubrimiento de hechos nuevos. Estas nuevas *teorías*, más “potentes”, reemplazan a las anteriores. De esta forma progresa el conocimiento *científico*: mediante desplazamientos progresivos de *teorías* científicas, partiendo de un *programa de investigación* (que se conceptúa como progresista o “degenerador”, dependiendo de si conduce al descubrimiento de hechos nuevos). “Las revoluciones científicas consisten en que un programa de investigación reemplaza a otro (superándolo de modo progresivo)” (Lakatos, 1974: 209). Aparece un nuevo esquema conceptual que sí proporciona una explicación veraz al problema existente.

Otro de los discípulos de Popper, Thomas Kuhn, propone el *relativismo sociohistórico* en la concepción de la ciencia, en su famosa obra *La estructura de las revoluciones científicas* (*The Structure of Scientific Revolutions*), publicada en 1962. En ella acuña el término *paradigma*, que especifica en 1969, coincidiendo con una edición en japonés de la obra citada:

“Por una parte, pasa por la completa constelación de creencias, valores y técnicas que son compartidas por los miembros de una comunidad determinada. Por la otra, denota una especie de elemento en tal constelación, las soluciones-enigmas concretos que empleados como modelos o ejemplos pueden reemplazar a reglas explícitas como base para la solución de los enigmas restantes de la *ciencia normal*” (Kuhn, 1978: 269).

La ciencia avanza no por *verificación* ni por *falsación* de las *teorías* existentes; sino por *sustitución* del modelo explicativo o *paradigma* vigente por otro posterior. Su aceptación dependerá de la existencia de consenso en la comunidad científica e inexistencia de evidencia que lo contradiga. Kuhn caracteriza la actividad *científica* como búsqueda de soluciones sustitutivas a los *paradigmas* en crisis. Las *revoluciones científicas* son las responsables del progreso del *conocimiento científico*.

Paul K. Feyerabend va más allá. Parte de que las *teorías* no pueden ser probadas ni refutadas de forma concluyente. Se muestra contrario al *método*, al actuar conforme a unas reglas metodológicas fijas. Defiende las *interpretaciones naturales* y el *conocimiento de sentido común*. Esto le lleva a propugnar la *contrainducción* (o la formación de *hipótesis* inconsistentes), y a encuadrarse dentro del *anarquismo metodológico*: “todas las metodologías tienen sus limitaciones y la única regla que queda en pie es la de que *todo vale*” (Feyerabend, 1974: 296). Planteamientos que comparte Derek Phillips, aunque discrepa, de manera radical, de los supuestos metodológicos de la ciencia *positiva*. Propone el abandono del *método* y de la comunidad científica que lo practica: “Intentemos mirar el mundo con nuestros propios ojos, y no a través de instrumentos científicos que con frecuencia nublan nuestra visión” (Phillips, 1973: 114).

En la actualidad, el *método científico* se muestra útil, no como ideal al que la práctica investigadora debiera aproximarse, sino como “referencia y patrón sobre el cual analizar los problemas y fases de la investigación en sociología” (Alvira, 1992: 62).

1.4.1. Integrantes esenciales del método científico

En 1971 Walter L. Wallace publica su libro *The Logic of Science in Sociology* (traducido al español en 1976), que ofrece uno de los esquemas más conocidos del proceso de investigación científica. Éste se representa como un círculo que conecta los cuatro componentes fundamentales del conocimiento (*teorías*, *hipótesis*, *observaciones* y *generalizaciones empíricas*) a través de cuatro *procesos cognitivos* diferentes (*deducción*, *operacionalización*, *interpretación* e *inducción*). Su esquema da lugar a otros posteriores (Bryman, 1988; Marshall y Rossman, 1989), que mantienen la concepción circular del proceso de conocimiento científico y de condicionamiento mutuo entre la *teoría* y la *investigación*.

La figura 1.1 sintetiza los componentes y procesos cognitivos presentes en la adquisición del conocimiento científico. Parte de la *teoría*, que tratará de comprobarse o verificarse. De la *teoría* se extraen (por medio de la *deducción lógica*) *hipótesis* específicas, que tendrán que operacionalizarse para proceder a su *contrastación* empírica. A la operacionalización de conceptos sigue la *observación* o recogida de datos empíricos. Éstos serán procesados para su posterior *análisis estadístico* o *semiológico*, dependiendo de la naturaleza de la información recabada y la metodología escogida en el *diseño* de la investigación. Sigue la fase de *interpretación* de la información y extracción de *generalizaciones empíricas*, que serán contrastadas con las *hipótesis* originales de la investigación. Si los datos empíricos contradicen las *hipótesis* del estudio, se procederá a su rechazo y a la consiguiente no aceptación de la *teoría* de la que se extrajeron las *hipótesis*. En caso contrario, las *hipótesis* serán aceptadas; pero ello no conllevará la *verificación teórica*, sino sólo su confirmación (en consonancia con el *falsacionismo* de Popper). Mediante el *método inductivo*, las

generalizaciones empíricas revierten en la *teoría*, que se verá confirmada, rechazada o ligeramente modificada. En caso de que no sea completamente apoyada por los *datos* empíricos, el proceso continúa. Las modificaciones introducidas en la *teoría* inicial de la investigación tratarán de comprobarse empíricamente. De este modo, la adquisición del *conocimiento científico* se configura como un *feed-back* de condicionamiento mutuo entre *teoría* y *observación*, a través de procesos de *deducción* (de la teoría a los datos, de lo general a lo concreto) e *inducción* (de los datos a la teoría, de los casos particulares a los principios generales).

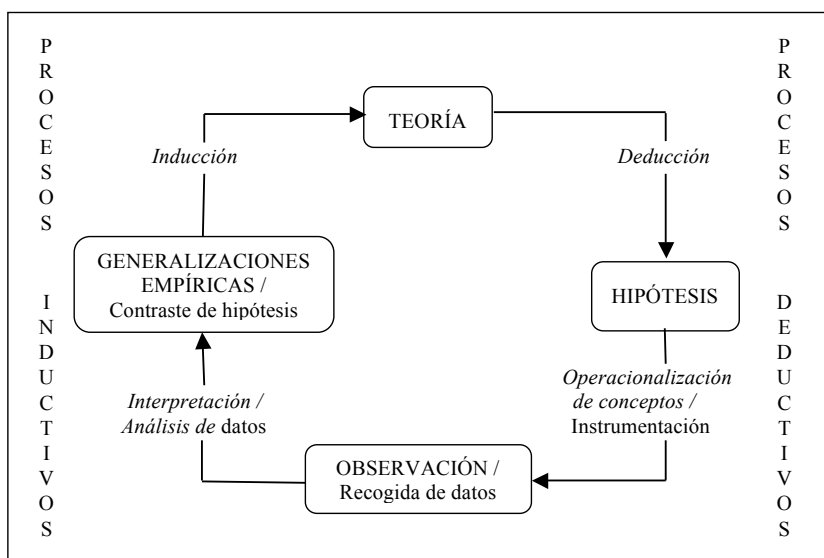


FIGURA 1.1. El proceso del conocimiento científico.

A) Teoría

“Un grupo de proposiciones lógicamente interconectadas, de las que pueden deducirse uniformidades empíricas” (Merton, 1983: 56). Está integrada por:

- *Conceptos*: “un símbolo lingüístico que categoriza a los fenómenos” (Phillips, 1985: 77) o “categorías para la organización de ideas y observaciones” (Bulmer, 1992: 43). “Sirven para clasificar y medir operativamente los fenómenos que se observan” (Blalock y Blalock, 1968: 7).
- *Proposiciones*: indican cómo y en qué condiciones generales se cree que se hallan relacionados los fenómenos definidos conceptualmente.
- *Axiomas* o *leyes*: explican las regularidades o conexiones esenciales que existen en los fenómenos sociales.

Para Wallace (1971) las *teorías* cumplen la función de *explicar* generalizaciones empíricas ya conocidas (subsumiéndolas en *conceptos* y *proposiciones*), y de *predecir* las aún desconocidas (guiando el futuro de la ciencia). Lazarsfeld (1973a) añade la función de *interpretación*: permite relacionar unos hechos empíricos con otros hipotéticos o ya verificados. La *teoría* además resulta imprescindible en las fases preliminares de la investigación, a la hora de:

- Decidir los *esquemas* clasificatorios a utilizar.
- Plantear los *conceptos* teóricos que orienten el análisis.
- Formular *problemas* de investigación que tengan una determinada relevancia social.
- Concretar *ideas* generales sobre cómo se producen los cambios sociales.
- Formular *hipótesis*: realizar previsiones a partir de descubrimientos aún no verificados.

Funciones de la *teoría* que se asemejan a las cinco diferenciadas por Merton (1967, 1983): formular *hipótesis*, encauzar el *problema* de investigación, analizar los *conceptos*, sugerir interpretaciones *post-factum* y *generalizar* los resultados de la investigación. Este autor diferencia las “grandes teorías” de las “teorías de rango medio”:

- Las *grandes teorías, teorías generales o globales*, ofrecen explicaciones bastante abstractas (especulativas y poco apoyadas en la evidencia empírica) de aspectos generales de la vida social. Por ello es difícil su materialización en una investigación empírica concreta. Si bien, algunos aspectos específicos de ellas pueden servir de base a la investigación guiando el proceso de investigación y proporcionando una explicación de los hallazgos.

Por ejemplo, la teoría marxista del desarrollo social y la teoría del sistema social de Parsons (1951). La primera proporciona una explicación genérica del desarrollo histórico de la sociedad moderna. La segunda vincula aspectos psicológicos y sociales de la conducta humana (en un intento de explicar las características definitorias de la sociedad occidental).

- Las *teorías de rango medio* se circunscriben a aspectos concretos de la vida social, a relaciones entre variables que puedan medirse de manera empírica. Su materialización en la investigación social es más fácil. De ellas se derivan hipótesis que pueden comprobarse empíricamente.

Por ejemplo, la teoría del suicidio de Durkheim. Tras observar que los católicos presentan tasas de suicidio inferiores a los protestantes, al igual que la población rural, los casados, las personas con hijos y las mujeres, trata de buscar una explicación a las regularidades que observa: ¿qué tienen en común las personas más propensas al suicidio? De sus observaciones deriva un principio más genérico:

las personas que se hallan menos integradas en la sociedad están más predispuestas al suicidio. De esta forma, las *generalizaciones empíricas* se convirtieron en significativas, o teóricamente relevantes, porque se pudieron relacionar o derivar de otra proposición o supuesto más abstracto: la cohesión social. La probabilidad de que esta *inducción* sea correcta se debe a que se analizaron diversas variables, obteniéndose el mismo resultado. Merton (1967: 151) extrae los siguientes supuestos que permitieron a Durkheim inferir la existencia de relación entre las variables tasa de suicidio y religión:

- La *cohesión social* proporciona apoyo psíquico a los miembros del grupo que están sujetos a ansiedades y tensiones fuertes.
- La tasa de *suicidio* es función de ansiedades y tensiones no resueltas.
- Los católicos presentan una mayor cohesión social que los protestantes, por lo que debería esperarse una tasa de suicidio inferior en la población católica que en la protestante.

B) Hipótesis

Se extraen de la *teoría* para su *contrastación* empírica. Se recomienda extraer más de una *hipótesis* para poder abarcar los diferentes aspectos que configuran la *teoría*.

“Deducir muchas hipótesis diferentes proporciona una comprobación más rigurosa del poder predictivo absoluto de la teoría, que derivar sólo una o unas cuantas hipótesis similares” (Brewer y Hunter, 1989: 37).

Las *hipótesis* representan *predicciones* o respuestas probables a los interrogantes que el investigador formula (ante un conocimiento previo) para su *contrastación* empírica. Suelen formularse como *proposiciones* en las que se afirma la existencia o inexistencia de relación esperada entre al menos dos *variables* (que se deducen de los *conceptos* y *proposiciones* que conforman la *teoría*). Además se hace referencia, aunque genérica, a las *unidades* de observación (que pueden ser individuos o agregados). En la investigación social, el individuo es la *unidad* de observación más habitual. De él se extrae información sobre sus características personales y su entorno sociocultural.

EJEMPLO DE EXTRACCIÓN DE HIPÓTESIS A PARTIR DE UNA TEORÍA

La *Teoría de la anomia* de Robert K. Merton (1938: “Social structure and anomie”. *American Sociological Review*, 3: 672-682), que tuvo varias revisiones y ampliaciones.

El término “anomia” fue introducido por Durkheim para designar aquella situación de carencia de normas que sigue a la desintegración del orden social. Merton lo reformula en

términos de “tensión” o desequilibrio entre metas y medios. Conceptúa la conducta desviada como descarga de la tensión psicosocial generada en el individuo ante el desequilibrio existente entre sus elevadas aspiraciones (producidas culturalmente) y los medios limitados para satisfacerlas. Un intento para reducir la tensión entre *metas* y *medios* es la conducta contranormativa. Los tipos de adaptación son: conformismo, innovación, ritualismo, retraimiento y rebelión. Del grado de tensión y del compromiso con las normas de la sociedad dependerá la implicación de la persona en conductas “desviadas”.

Axiomas (o aseveraciones) principales de esta teoría son:

1. Las contradicciones entre la estructura cultural (metas) y la social (medios institucionalizados) produce una tendencia a la *anomia*.
2. Las sociedades “igualitarias” (tipo sociedad norteamericana de la época) incitan a todos los individuos a alcanzar las mismas metas (éxito económico), sin facilitarles igual acceso a los medios legítimos para alcanzarlas.
3. La “tensión” se genera principalmente en los miembros menos protegidos, aquellos con poca educación y escasos recursos monetarios.
4. Ésta se resolverá dependiendo del grado de tensión y del compromiso del individuo con las normas de la sociedad.

De estos comienzos axiológicos se deducen *proposiciones* de las que derivan *hipótesis* como:

- La comisión de actos delictivos es mayor entre las personas de clase social baja.
- Cuanto menos comprometida esté la persona con las normas de la sociedad, más predispuesta estará a su transgresión.

C) Observación

Una vez seleccionado el ámbito de la realidad social a analizar, se procede a la *observación* o recogida de información. El investigador recoge datos relacionados con los *indicadores* empíricos que ha seleccionado para su posterior análisis, en conformidad con la metodología elegida en el *diseño* de la investigación. Bien mediante *cuestionario autoadministrado*, *entrevista personal*, utilizando *datos disponibles* (informes policiales, judiciales, escolares), o cualquier otra técnica de recogida de datos que el investigador estime necesaria; o bien una conjunción de diversas técnicas, como ya se vio en el apartado 1.3.

D) Generalizaciones empíricas

El círculo del proceso de investigación científica se cierra cuando los resultados de la investigación revierten (mediante la operación metodológica de la *inducción*) en la *teoría* inicial. La información que se obtuvo en la fase de *observación* se interpre-

ta en busca de *generalizaciones empíricas*: “cualquier proposición aislada que resuma uniformidades observadas de relaciones entre dos o más variables” (Merton, 1983: 171). Las *generalizaciones empíricas* se contrastan con las *hipótesis* de la investigación. Si los datos empíricos corroboran las *hipótesis* teóricas, la *teoría* se confirma. En caso contrario, la *teoría* queda debilitada por los hallazgos empíricos de la investigación. Lo que conlleva su modificación o rechazo.

“Rara vez la teoría inicial es completamente apoyada por la investigación. Esto es bueno porque nos hace pensar, modificar o desarrollar la teoría inicial y es lo que nos lleva al progreso. Como tal la teoría modificada necesitará ser comprobada rigurosamente” (De Vaus, 1990: 19-20).

El proceso de investigación comienza de nuevo. Si bien, téngase presente que “cualquiera que sea el método utilizado en la contrastación de hipótesis, los resultados nunca son ciertos, sino aproximaciones en términos de *probabilidad*” (Goode y Hatt, 1952: 87). Por esta razón, para cada *teoría* se tratará de “especificar los hechos que la confirman y la probabilidad de la teoría a la luz de estos hechos” (Lakatos, 1975: 208).

Merton (1967, 1983) resume en tres las formas principales en que la *investigación* empírica ayuda a la *formulación* de la *teoría*:

- Lo que denomina “*serendipity pattern*”: aquella situación en la cual hallazgos de investigación no anticipados (que no son consistentes con la *teoría* prevaleciente u otros hechos establecidos) estimula el desarrollo de una *nueva teoría* o la modificación de la *teoría* existente.
- La *observación* repetida de hechos (que anteriormente se habían negado) presiona hacia la modificación o ampliación de la *teoría*.
- El proceso de investigación fuerza al investigador a establecer *definiciones operacionales* de *conceptos*. Esto ayudará a la clarificación de la *teoría* o *marco teórico*.

Merton es uno de los autores que más insiste en la necesidad de que los investigadores desarrollen una orientación teórica: realicen investigaciones en el contexto de un *marco teórico*, comprobando *teorías* y modificándolas a la luz de los resultados de la *investigación*. La *teoría* y la *investigación* han de estar interrelacionadas. Toda *teoría* ha de basarse en *hallazgos empíricos* para que no quede en una mera especulación ideológica. De igual manera, toda *investigación* debe partir de una *teoría* para adquirir una mayor relevancia. Pero no siempre se ha reconocido esta necesidad mutua de la *teoría* y la *investigación*.

Por último, conviene puntualizar que no todas las *investigaciones* empíricas cumplen cada una de las fases enunciadas, y de forma lineal. Depende del estudio (si se fundamenta en *datos primarios*, *secundarios* o *terciarios*) y de la *metodología* seguida (*cuantitativa* o *cualitativa*). Ambas *metodologías* comparten unas fases similares en su desarrollo básico: selección y definición del *problema* a partir de una *teo-*

ría o investigación previa, formulación de *hipótesis*, elección del *diseño* de investigación, recogida de *datos*, *análisis* de la información, *validación* e *informe*. Si bien, la metodología *cualitativa* se caracteriza por una redefinición constante, tanto del *objeto* de estudio como de las *hipótesis* de investigación, hasta lograr su adecuación. De ahí que las diferentes fases del proceso de *conocimiento científico* no acontezcan de manera gradual, sino interactivamente. El análisis de los datos empíricos puede implicar la modificación de las *hipótesis* de partida, con la consiguiente recogida de información y, así, sucesivamente. Lo mismo acontece en la investigación *cuantitativa*. En ella tampoco se sigue siempre, y de forma rígida, el esquema señalado. En suma, la práctica de la investigación social no siempre sigue una sucesión ordenada de fases. Más bien se trata de un *feed-back* continuo.

1.4.2. Distinción entre proyecto y diseño de investigación

Una vez definido el método científico y sus condicionantes, corresponde pasar del proceso de investigación “ideal” a la práctica habitual. Ésta se halla determinada por los *objetivos* y el *marco teórico* del estudio. Pero también intervienen aspectos *económicos* (los *recursos* a disposición del investigador) y *temporales* (el plazo de *tiempo* concedido para su materialización). La figura 1.2 sintetiza las diferentes fases y componentes de un *proyecto* de investigación. Abarca el *diseño* de la investigación: un “plan global de la investigación que integra, de modo coherente y adecuadamente correcto, técnicas de recogida de datos a utilizar, análisis previstos y objetivos” Alvira (1989: 85); o una secuencia lógica que relaciona los datos empíricos con las preguntas formuladas al inicio de la investigación y las conclusiones. Es “más que un plan de trabajo”, como señala Yin (1989: 29).

Además del *diseño*, el *proyecto* incluye tres componentes esenciales: la formulación del *problema* de investigación, su *operacionalización* y la *factibilidad* de la investigación. En esta última se concreta el detalle económico, organizativo y temporal de las diferentes fases del proceso de investigación. Cada uno de los cuatro componentes del proyecto se verá afectado por tres elementos claves: los *objetivos* de la investigación, los *recursos* (o medios materiales, económicos y humanos de que disponga el investigador para hacer viable el estudio) y el *tiempo* concedido para su realización. Su conjunción marcará el alcance y el desarrollo de la investigación, por su incidencia directa en el *diseño* de la investigación.

A) Formulación del problema de investigación

Como ya afirmase Simon (1978: 98), “No importa en qué problema quieras trabajar y no importa qué método eventualmente utilizarás, tu trabajo empírico deberá comenzar con una consideración cuidadosa del problema de investigación”. Ha de concretarse:

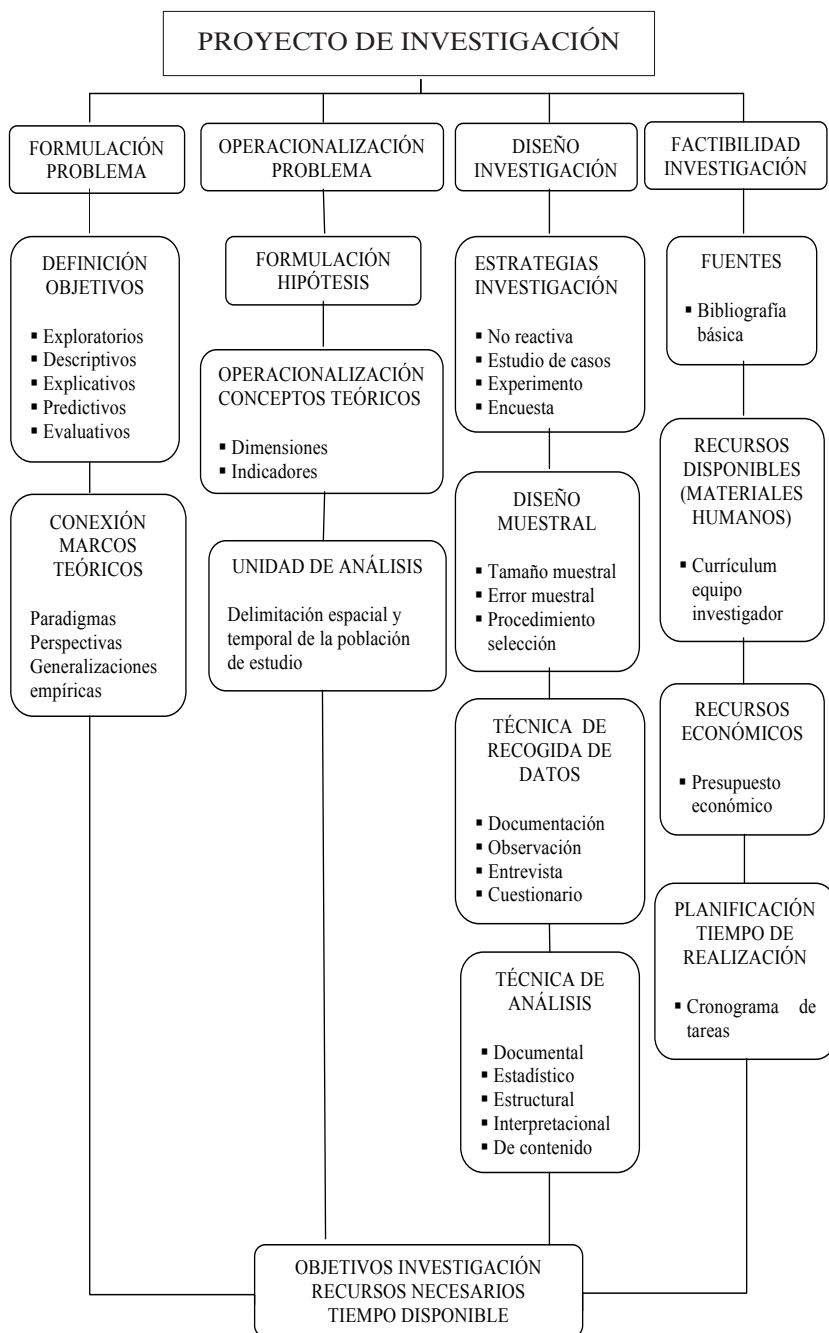


FIGURA 1.2. El proyecto de investigación.

- Qué es lo que se pretende analizar: los *objetivos* generales y los específicos de la investigación.
- Qué *razones* motivaron la elección de ese *problema*.
- *Justificar* dicha elección, enfatizando la relevancia del tema y su trascendencia. Ha de hacerse de forma convincente para lograr la aprobación del *proyecto* de investigación y consiguiente financiación económica (en caso de solicitarse).

Al principio, la *idea* o *problema de investigación* se presenta de forma vaga y genérica. Paulatinamente irá concretándose, a medida que el investigador va ahondando en ella. La *fuerza* originaria de la *idea* puede ser muy dispar. A veces es la propia entidad o institución que financia la investigación la que propone el *problema* o tema de estudio. En otras ocasiones, es el azar el que suscita la *idea*: en el curso de una conversación; durante la lectura de un libro o artículo de prensa (periódico, revista); o tras la observación de un acontecimiento, de forma directa o indirecta (por medio de la televisión, la radio). Pero la *idea* puede también ser el resultado de una reflexión profunda, presente durante mucho tiempo en la mente del investigador, aunque hasta entonces no haya sido concretada ni operacionalizada en un *proyecto* de investigación. Al principio es sólo una *idea*. El investigador tendrá que irle dando forma y concreción. Para ello deberá adentrarse en el campo de conocimiento determinado donde se ubique la *idea*. A tal fin se recomienda:

- a) Realizar una *revisión bibliográfica* exhaustiva sobre el *tema* de interés y otros similares que puedan orientar en su indagación. En especial, cuando apenas se encuentra bibliografía teórica y empírica sobre el tema, por su novedad o escasa repercusión en la tradición investigadora anterior. El objetivo es familiarizarse con el *tema* escogido, sintetizando y evaluando lo analizado hasta entonces. De la *revisión* se extraerá información sobre:
 - La *metodología* seguida y los *resultados* obtenidos con ella (sus potencialidades y límites). Ello ayudará al *diseño* de la investigación.
 - Aspectos del *problema* tratados, aquellos no abordados y los suscitados tras la conclusión de la investigación (relevantes para una futura indagación).

Se trata de discernir si la *teoría* existente y la *investigación* previa sugieren una respuesta (aunque sea parcial) a las *preguntas* de investigación, o una dirección a seguir dentro del estudio. La *revisión bibliográfica* contribuye a:

- *Estructurar la idea* originaria del proyecto, contextualizándola en una perspectiva teórica y metodológica determinada.
- *Prevenir errores* cometidos en estudios previos.
- *No ser redundantes*, dando a la nueva investigación un enfoque diferente del tema o aplicar una estrategia de indagación distinta.

- *Plantear otros interrogantes* de investigación para cubrir aspectos menos observados (y no reincidir en aquellos continuamente analizados).

b) *Indagaciones exploratorias*, que pueden materializarse en:

- *Discutir el tema con otros investigadores* que puedan aportar ideas de cómo llevar a cabo la investigación y qué pasos seguir.
- *Entrevistar a algunos de los implicados* en el *problema* a investigar, con el propósito de recabar, de ellos, información que ayude al *diseño* de la investigación.

Ambas tareas previas (desde la *revisión bibliográfica* a las *entrevistas abiertas*) son cruciales, cuanto más vaga sea la *idea* originaria del estudio y menos conocimiento se tenga del área donde se ubique. A partir de ellas se configurará el *marco teórico* de la investigación, que además le proporcionará un marco de referencia al que acudir para la interpretación de los resultados que recabe la investigación. A veces se partirá de una *teoría*; otras veces, el *problema* de estudio se imbricará en varias *perspectivas teóricas* que cubran distintos aspectos del *problema* a analizar. Pero, en más ocasiones de las deseables, el *marco teórico* se fundamenta en *generalizaciones empíricas* (o *proposiciones* comprobadas en investigaciones previas que no han logrado cristalizar en una *perspectiva teórica* concreta). De lo exhaustivo que se sea en esta fase previa del *proyecto* dependerá el buen desarrollo de la investigación.

B) *Operacionalización del problema de investigación*

A la definición y delimitación del *objeto* de estudio sigue su operacionalización, tanto en la vertiente *teórico-analítica* (conceptos, categorías analíticas, variables a analizar) como *poblacional* (población a estudiar y su delimitación temporal y espacial). Incluye la formulación de *hipótesis*, que se deducen del *marco teórico* de la investigación. Como se viera en el subapartado 1.4.1, las *hipótesis* representan respuestas probables que el investigador avanza a las *preguntas* iniciales de la investigación. Se expresan en forma de proposiciones, en las que se afirma la existencia o inexistencia de asociación al menos entre dos variables (dependiente e independiente, generalmente), en una situación determinada. Se recomienda que las *hipótesis* se encuentren relacionadas con los *objetivos* de la investigación y sean lo más precisas posible.

Por ejemplo, el consumo de tabaco disminuye la esperanza de vida de los jóvenes. O, el interés por la asignatura aumenta la atención del alumno en clase.

Las *hipótesis* contienen *conceptos* que determinan lo que ha de analizarse. Éstos han de *operacionalizarse*; es decir, traducirse en dimensiones e indicadores que per-

mitan su medición, como se verá en el capítulo 2. Además suelen incluir a la *unidad de análisis*: la *población* (o universo objeto de estudio). Que se elija una u otra *población* puede tener implicaciones significativas en el contenido de la investigación y en sus costes (económicos y de tiempo). De ahí que Hakim (1994) recomiende que la *población* se establezca en los *objetivos* de la investigación, al determinar la información a extraer y la naturaleza de los resultados de la investigación. Pero, las *unidades de análisis* no sólo incluyen el *ente* individual o colectivo que será objeto de observación. También, el *espacio* donde se ubica y el *tiempo*. En los estudios *longitudinales* el *tiempo* figura (explícitamente) como una de las *variables* claves de la investigación.

C) Diseño de la investigación

A la *operacionalización* del problema sigue la concreción de la *estrategia* a seguir para alcanzar los *objetivos* de la investigación. Al igual que la fases precedentes, la configuración del *diseño* de la investigación estará determinada por los *objetivos* del estudio (exploratorios, explicativos, predictivos, evaluativos), los *recursos* (materiales, económicos, humanos) y el plazo de *tiempo* disponible para su materialización. Por lo que, el investigador ha de tener en mente tanto los *objetivos* como la *factibilidad* del *diseño*. Ha de procurar que sea viable, con los recursos disponibles y dentro de los límites de tiempo fijados.

Esta triple consideración de *objetivos*, *recursos* y *tiempo* incide directamente en la elección de la *estrategia* de investigación: la investigación *no reactiva*, el *estudio de casos*, la *encuesta*, el *experimento*. Su uso exclusivo, combinado o articulado con otros métodos de investigación. La *estrategia* elegida influirá en:

- a) El *diseño muestral*: el tamaño de la *muestra* (unidades de observación) y el procedimiento de selección.
- b) Las *técnicas de recogida de información* (de documentación, observación y entrevista). En concreto:
 - *Revisión de fuentes de observación secundaria* (estadísticas, documentos).
 - *Observación sistemática, participante y no participante*.
 - *Entrevistas abiertas, semi o no estructuradas* (individuales o grupales).
 - *Relatos biográficos* (múltiples, paralelos, cruzados) y *documentos personales*.
 - *Cuestionario estandarizado* (sondeos en serie, tipo panel, macroencuestas).
- c) Las *técnicas de análisis de datos*:
 - *Documental*.
 - *Estadístico* (univariable, bivariable, multivariable).

- *Estructural* (análisis del discurso, etnografía, fenomenografía).
- *Interpretacional* (construcción de teorías, descriptivo/interpretativo).
- *De contenido* (cuantitativo y cualitativo).

Las diferentes *técnicas cuantitativas* de recogida y análisis de datos se abordarán en los capítulos que siguen. Las correspondientes a la metodología *cualitativa* se tratan en manuales especializados, como el publicado en esta misma editorial por Valles Martínez (1997).

En la elección de la *estrategia* y de las *técnicas* de recogida y análisis de la información también incide:

- a) Si la investigación se ceñirá a un momento temporal concreto (*diseños seccionales o transversales*) o incluirá diferentes períodos de tiempo (*diseños longitudinales*).
- b) La *formación metodológica* del investigador.
- c) El grado de *rigor y precisión* que desee que tenga su indagación.
- d) La *factibilidad* de la investigación: los recursos materiales y económicos disponibles en la franja temporal que se dispone para su materialización.

Pero la práctica real de la investigación es muy dispar. De acuerdo con Ibáñez (1989), el investigador social suele elegir (sin pensar demasiado en la elección) la *técnica* que tiene más a mano, bien por razones *personales* (es experto en esa técnica), *organizativas* (trabaja en una organización que trabaja con esa técnica) o *institucionales* (pertenece a una institución interesada en vender esa técnica). Sea como fuere, el investigador deberá justificar (en su *proyecto* de investigación) por qué eligió unas *estrategias* y *técnicas* con preferencia sobre otras. La exposición de las razones suele fundamentarse en tres aspectos esenciales: los *objetivos*, las *características* y las *condiciones de realización* de la investigación.

La puesta en práctica de la investigación puede llevar a modificar el *diseño* inicial. Por ello conviene prever alternativas al *diseño* original de la investigación.

D) *Factibilidad de la investigación*

El proyecto concluye con la exposición de las condiciones mínimas requeridas para hacer viable (o factible) la investigación diseñada. Incluye:

- a) *Fuentes bibliográficas*. En el proyecto suele adjuntarse una *bibliografía* básica, especializada en el tema que constituye el eje de la investigación, tanto en su vertiente teórica como en la empírica y metodológica. Consistirá en una selección de *obras claves* y de publicaciones actualizadas.
- b) *Recursos* (materiales y humanos) *disponibles*. El equipo investigador acredita su *capacidad* (experiencia) y *medios* de los que dispone para la realización

del proyecto. Al proyecto se añade el currículum vitae del investigador principal y de los integrantes del equipo investigador. En especial se resalta la *experiencia* tenida en el área de conocimiento donde se inserta el *problema* o *tema de investigación*.

c) *Otros recursos* (materiales, económicos y humanos) *necesarios*. Si se solicita financiación económica, en el proyecto se especifica la cantidad que se precisa y se desglosa en diferentes partidas:

- *Personal* (que habría que contratar con cargo al proyecto).
- *Equipo material inventariable*: ordenadores, impresora, cámaras de vídeo... (pequeño equipamiento científico-técnico más material bibliográfico).
- *Material fungible*: material de papelería y oficina, fotocopias, teléfono, correos.
- *Viajes y dietas*, más otras partidas que se precisen.

d) *Planificación del tiempo de ejecución de la investigación*. El investigador concreta la *cronología de tareas*. Delimita la duración de cada fase de la investigación (los meses que ocupará, junto al personal responsable de cada tarea). Además especifica los *plazos de entrega parciales* (si los hubiere) y la *entrega final*.

Existen diferentes técnicas para calcular los tiempos de duración de cada etapa del proyecto, su coste y las posibles variaciones que puedan acontecer. Una de las más populares es el *método PERT* (*Program Evaluation and Review Technique* o *Program Evaluation/Research Task*): un *grafo* en el que se especifica cada una de las *tareas*, cómo dependen unas de otras, en qué *tiempo* se realizarán y qué *probabilidad de retraso* existe en su ejecución. Para su cálculo se pregunta a los responsables de cada una de las tareas el tiempo de duración habitual de su trabajo.

Todos estos detalles que conforman la última fase del *proyecto* de investigación son esenciales, en cuanto que ayudan al buen desarrollo de la investigación proyectada. Por lo que no hay que minusvalorar el esfuerzo que se dedique a su concreción. Como bien apunta Hakim (1994: 157), “quizás el error práctico más común consiste en subestimar el presupuesto, en tiempo y dinero, requerido para un proyecto”.

Por último, reiterar que se trata de un “proyecto” cuya puesta en práctica puede llevar a modificar una o varias de las partes que lo forman. De la pericia e ingenio del investigador, además de los medios de los que disponga, dependerá que la investigación se cumpla con éxito.

EJEMPLO DE PROYECTO DE INVESTIGACIÓN

El proyecto *MEXEES II* (CSO2009-07295), dirigido por M.^a Á. Cea D’Ancona, y aprobado en la convocatoria de ayudas de *Proyectos de Investigación Fundamental no orientada* del Ministerio de Ciencia e Innovación en 2009.

En el proyecto figura, en primer lugar, su *título*: “La medición de la xenofobia en España (II): desarrollo de modelos explicativos-comprensivos para la implementación de políticas de integración social”. Después, un *resumen* donde se destaca, de forma breve y precisa, los aspectos más relevantes y los objetivos propuestos. Se señala que es la continuación de un proyecto anterior (SEJ2005-00568), pero que se quiere dar un paso más hacia la comprensión de la xenofobia *versus* xenofilia: cubrir interrogantes que suscitan los actuales escenarios sociales y políticos (marcados por la crisis económica), además de constatar hallazgos de la anterior investigación.

El *objetivo* principal de la investigación es avanzar en el desarrollo de modelos explicativos-comprensivos de las distintas modalidades de racismo y xenofobia que confluyen en la España actual. Se busca la mejora de su medición y comprensión, que ayude a la propuesta de políticas encaminadas a su erradicación o aminoración.

Dicho objetivo se concreta en cuatro *objetivos específicos*:

1. La generación de *teoría* vía la *conceptualización* y *mapeado* de relaciones entre etnicidad, identidad, biografía y estructura social.
2. La determinación de *dimensiones* e *indicadores* precisos de xenofobia (“manifiesta” y “latente”).
3. La profundización en los discursos xenófobos y xenofílicos.
4. Ahondar en el *estudio de casos* que generan más xenofobia y xenofilia para progresar en la sociogénesis de las modalidades de xenofobia.

Dos son las *hipótesis* claves que quieren constarse:

1. El temor a la competencia se antepone al miedo a la pérdida de la identidad nacional-cultural como desencadenante de xenofobia en contextos de déficit de recursos.
2. La crisis económica propicia la exteriorización de la xenofobia “latente” incluso en indicadores directos, más expuestos al sesgo de deseabilidad social.

En la *introducción* (máximo cinco páginas) se describe la finalidad del proyecto, los antecedentes y el estado actual de los conocimientos científico-técnicos, incluyendo la bibliografía más relevante, los grupos nacionales o internacionales que trabajan en la materia específica del proyecto o en materias afines. Sigue una descripción más pormenorizada de los *objetivos* e *hipótesis* del proyecto, junto con los antecedentes y resultados previos, del equipo solicitante o de otros, que avalan la validez de la hipótesis de partida (en un máximo de dos páginas). De los *objetivos* se pide que se enumeren brevemente, pero con claridad, precisión y de manera realista (es decir, acorde con la duración prevista del proyecto). También, que se destaque la novedad y la relevancia de los *objetivos* (porque se mencionan explícitamente en los criterios de evaluación de las solicitudes).

El siguiente apartado corresponde a la *metodología y plan de trabajo*. Se resalta que se detalle y justifique con precisión la metodología y el plan de trabajo que se proponen, junto con la planificación temporal de las actividades, incluyendo *cronograma*. Se desglosa cada una de las tres fases del proyecto, en qué consiste, su justificación, qué miembros del equipo participarán y cuánto tiempo durará cada fase:

1. El uso de fuentes documentales y estadísticas (11 meses).
2. Investigación cualitativa mediante grupos de discusión, entrevistas biográfico-narrativas y técnicas etnográficas de observación-participación con población autóctona y extranjera en contextos de la vida cotidiana (12 meses).
3. La encuesta y redacción final del informe (13 meses).

Sigue el apartado correspondiente a los *beneficios* del proyecto, la difusión y explotación de los resultados (máximo una página). Han de destacarse las contribuciones científico-técnicas esperables del proyecto y los beneficios para el avance del conocimiento y de la tecnología. El plan de difusión y de explotación de los resultados del proyecto se valora en el proceso de evaluación de la propuesta y en el de seguimiento del proyecto.

Aunque junto al proyecto se adjuntan los currículos de cada uno de los miembros del equipo investigador, en la *Memoria* del proyecto hay un apartado específico sobre el *Historial* del equipo solicitante en el tema propuesto. En un máximo de dos páginas hay que indicar las actividades previas del equipo y los logros alcanzados en el tema propuesto. La finalidad de este apartado es determinar la adecuación y capacidad del equipo en el tema, además de la viabilidad de la actividad propuesta. A él se suma el detalle de los proyectos y contratos I+D de cada miembro del equipo investigador en los últimos cinco años.

En la *Solicitud* del proyecto se incluye el detalle correspondiente al *Presupuesto* que se solicita, su descripción, coste y justificación, en cada uno de los apartados correspondiente a:

- Costes del *personal* contratado con cargo al proyecto.
- Costes de *ejecución* correspondientes a:
 - Pequeño equipamiento científico-técnico y material bibliográfico.
 - Material fungible.
 - Viajes y dietas.
 - Varios.

1.4.3. Tipologías de diseños de investigación

Existe una amplia tipología de *diseños* de investigación en función de los *objetivos* del estudio, el *marco temporal* (cómo se aborde la variable tiempo) y el *contextual* de la observación (entorno natural, semiartificial o artificial). El cuadro 1.3 resume tres clasificaciones principales en la metodología cuantitativa.

CUADRO 1.3. Clasificación de los diseños de investigación

CRITERIO DE CLASIFICACIÓN	DISEÑO DE INVESTIGACIÓN
<i>Objetivos de la investigación</i>	Exploratorios Descriptivos Explicativos Predictivos Evaluativos – De impacto – Del proceso – De necesidades – Mediante análisis de sistemas – Coste-beneficio – De conjunto
<i>Tratamiento de la variable de tiempo</i>	Seccionales o transversales Longitudinales – De tendencias – De cohorte – Panel
<i>Cumplimiento de los supuestos de la experimentación</i>	Preexperimentales o correlacionales Cuasiexperimentales Experimentales

A) Diseños exploratorios, descriptivos, explicativos, predictivos y evaluativos

Una primera tipología de *diseños* de investigación responde al *objetivo* principal de la investigación. Aunque una investigación puede incluir varios objetivos, se diferencia entre:

a) *Diseños exploratorios*: cuando la investigación se realiza para:

- Familiarizarse con el *problema* de investigación y conocer qué aspectos precisan un análisis más pormenorizado en indagaciones posteriores.
- Verificar la *factibilidad* de la investigación y documentar los medios necesarios para hacerla viable.
- Comprobar qué *estrategia* de investigación se adecua más a su análisis.

Rara vez el estudio *exploratorio* constituye un fin en sí mismo. Pero sí es el paso a otros posteriores. A ello contribuye la escasa representatividad de

los casos que se analizan (ya sea por su volumen o por el procedimiento de selección muestral utilizado).

- b) *Diseño descriptivo*: puede ser un fin en sí mismo, o un paso previo en el proceso de investigación (al igual que el *exploratorio*). La explicación de cualquier realidad o evento precisa de su previa descripción mediante alguna o varias *estrategias* de investigación (*encuesta, uso de documentos y estadísticas* u otra modalidad *no reactiva, experimentación, estudio de casos*). De ellas, el investigador recabará información que le servirá para caracterizar su objeto de estudio.
- c) *Diseño explicativo*: a la descripción sigue la búsqueda de causas o razones de los hechos, acciones, opiniones o cualquier fenómeno que se analice.

Por ejemplo, tras describir las cifras de la inmigración (el número y perfil de los inmigrantes extranjeros), a partir de datos estadísticos disponibles, avanzar en el conocimiento de las causas de la inmigración mediante un estudio diseñado al efecto (a través de una encuesta a la población extranjera, un estudio de casos...).

- d) *Diseño predictivo*: cuando el objetivo del estudio es “predecir” la evolución del fenómeno que se analiza. La *predicción* constituye un objetivo en sí mismo, aunque a su vez conlleva uno o varios de los objetivos anteriores (la *descripción* y *explicación* de lo que se analiza).
- e) *Diseño evaluativo*: trata de “evaluar” la efectividad de un determinado programa (o conjunto de actividades específicas). La evaluación puede adoptar diferentes modalidades:

- *Impacto*: analiza los efectos del *programa* en busca de su mejora. Es una de las modalidades de evaluación más populares. Precisa que previamente se definan los criterios de éxito (objetivos o metas que el programa debería alcanzar), se diferencien los resultados que se deban al programa (de los causados por otros factores), y se especifiquen las condiciones bajo las cuales el programa resultaría más eficaz.
- *Proceso, formativa* o del *desarrollo*: su objetivo fundamental es la descripción del *programa*, en qué consiste y cómo funciona en la teoría y en la práctica. De ahí su complementariedad con la modalidad de evaluación anterior. Describe si el programa se está realizando de acuerdo a como fue configurado. Además permite conocer las causas de su éxito o fracaso.

Por ejemplo, evaluar la efectividad de la justicia de menores (si logra frenar la carrera delictiva del menor) requiere la evaluación previa del *proceso* seguido en la aplicación de la norma legal y de las actuaciones correctivas desarrolladas.

- *Valoración de necesidades*: tiene por objetivo la identificación de las necesidades prioritarias que el programa debe atender.

- *Mediante análisis de sistemas*: el programa se conceptúa como un conjunto de subsistemas (personal subalterno, dirección, gerencia) relacionados y vinculados, a su vez, a otros sistemas más amplios (familia, comunidad). Se estudian sus relaciones e interrelaciones.
- *Análisis coste-beneficio*: se analizan los costes del programa con sus resultados (generalmente en términos monetarios). Difiere del *análisis efectividad-coste* en que éste último no traduce los beneficios del programa a términos monetarios, sino de *impacto* real.
- *De conjunto*: comprende tanto las metas del programa como su repercusión, las alternativas disponibles y los costes de su actuación. Representa la síntesis de los anteriores.

La interpretación de los *resultados* de cualquier *programa*, intervención, o conjunto de actividades, requiere que antes se haya descrito el *programa* y su funcionamiento real. A partir de esta descripción se procede a descubrir qué se debe al *programa* y qué cabe atribuir a la intervención de otros factores.

“Evaluar es más que investigar (...) Evaluar es emitir juicios de valor, adjudicar valor o mérito de un programa/intervención, basándose en la información empírica recogida sistemática y rigurosamente” (Alvira, 1991a: 7, 10-11).

B) Diseños seccionales (o transversales) y longitudinales

Una clasificación distinta de los diseños de investigación responde a cómo se haya planificado la recogida de información: en un único momento (diseños *seccionales o transversales*) o en dos o más veces (diseños *longitudinales*) para analizar la evolución de los fenómenos que se analizan a lo largo del “tiempo”. El “tiempo” pasa a ser variable clave en esta clasificación de diseños.

- a) *Diseños seccionales o transversales*: la recogida de información se produce en un único momento en el tiempo, aunque se recojan datos de diferentes entornos y épocas. El objetivo de la investigación puede ser *descriptivo* (describir las características de una población en una fecha concreta, caso del censo de población, *por ejemplo*) o *explicativo* (analizar las variables que inciden en la ocurrencia de un hecho o acción determinada, mediante una encuesta *por ejemplo*).

Por ejemplo, una *encuesta* realizada (al mismo tiempo) en los distintos países de la Unión Europea para conocer las actitudes de los europeos ante las minorías étnicas. El estudio sería *longitudinal* si la misma encuesta se aplicara en diferentes años para analizar la evolución de las actitudes a lo largo del tiempo.

b) *Diseños longitudinales*: se planifica la recogida de información en diferentes momentos temporales con el objetivo de analizar su dinámica. Las fechas de recogida se especifican en el proyecto de investigación. La amplitud del período de observación y la cronología se halla relacionada con los objetivos del estudio. Se distingue entre:

- *Diseño longitudinal de tendencias*: se analiza la evolución de la población en distintos momentos en el tiempo ya con datos estadísticos o demoscópicos.

Por ejemplo, la serie de informes sobre la evolución del racismo y la xenofobia en España que el Observatorio de Racismo y Xenofobia (OBERAXE) realiza cada año, desde 2008, aplicando la misma encuesta y diseño muestral (tamaño y procedimiento de selección). Pero los sujetos de los que se extrae información demoscópica difieren cada año.

La existencia de bancos de datos estadísticos y de encuesta facilita el análisis longitudinal de *tendencias* como se verá en el capítulo 4.

- *Diseño longitudinal de cohorte*: difiere del anterior en que el interés no se halla en la población total, sino en una subpoblación o *cohorte*. La *cohorte* la integran individuos que comparten una misma característica, como la edad. La *cohorte* de edad puede definirse de forma amplia (personas nacidas en la década de los noventa, *por ejemplo*) o más específica (personas que nacieron en 1992). Pero la *cohorte* puede también determinarse a partir de un acontecimiento (personas que se licenciaron en 2008 o parejas que se casaron en 2010) o experiencia concreta (jóvenes en paro desde 2010 o personas que estuvieron en paro durante 10 años, *por ejemplo*). Se quiere analizar su evolución. Para ello se selecciona una *muestra* distinta de entre las personas que integran la *cohorte*, en cada fecha de recogida de información.

Por ejemplo, se desea conocer la trayectoria laboral de los licenciados en sociología. La *cohorte* fijada la componen los licenciados en universidades públicas y privadas en España en 2009. Cada dos años se extrae una *muestra* de entre los licenciados en dicha fecha. Si la *cohorte* se fija en la edad de 30 años, cada dos años se extraerá una *muestra* de entre los licenciados que tengan 30 años. A diferencia del supuesto anterior, la edad permanece estable; en el anterior la edad aumenta cada dos años (con la consiguiente merma en la base muestral).

Pero la investigación no tiene por qué circunscribirse a una única *cohorte*. Puede compararse la evolución de *cohortes* diferentes.

Por ejemplo, alumnos que se licenciaron en los cursos 1997-1998 y 2008-2009 para analizar su trayectoria laboral y los efectos debidos a haber cursado planes de estudios diferentes.

- *Diseño longitudinal de panel*: se analiza la evolución de unos mismos individuos (que se eligen al inicio de la investigación) a lo largo del tiempo. A diferencia de los anteriores, no se procede a nuevas selecciones muestrales para cada momento en que se recoge información.

Por ejemplo, si la encuesta se pasase a la misma *muestra* de licenciados en sociología en 2009 cada dos años (para conocer la trayectoria de las mismas personas a lo largo del tiempo; no de una generación concreta, determinada por una cohorte, ni de los jóvenes en general o personas de una misma edad en cada fecha).

El *diseño panel* es de gran utilidad para indagar en las causas del “cambio”. Las personas de las que se extrae información son las mismas. Lo que varía son las circunstancias en que se hallan tras el paso del tiempo. Ello permite analizar los factores que pueden haber contribuido al cambio que en ellos se observe. En su contra están dos problemas metodológicos fundamentales:

- El *desgaste de la muestra* con el consiguiente aumento de la *no respuesta*. Un problema similar al denominado “mortalidad experimental”. Las personas que inicialmente componen la *muestra* del estudio van mermándose con el paso del tiempo, ya sea porque se nieguen a seguir participando en la investigación, o porque no se pueda contactar con ellas (cambio de domicilio, enfermedad, fallecimiento...). Su efecto es mayor cuanto más se alargue el estudio en el tiempo. El tamaño de la *muestra* se reduce y, además, puede introducir un sesgo importante en la fase final de la investigación, si quienes abandonan el estudio presentan unas características sociodemográficas diferentes a aquellos que siguen participando en él (repercute en la *validez externa* de la investigación).
- Efectuar varias mediciones en la misma población en fechas diferentes puede suscitar *sesgos* debidos al aprendizaje. Estas acontecerán con más probabilidad cuanto más próximas estén las fechas de recogida de información.

Estos dos problemas se suman a uno común a los diseños *longitudinales*: el mayor coste económico de la investigación, de mayor alcance conforme aumenta el tamaño de la *muestra* y su diversidad espacial.

Los *diseños transversales o seccionales* también permiten el análisis del cambio (y a un menor coste) si en el instrumento de medición (cuestionario, guion de entrevista abierta...) se incluyen preguntas relativas al “pasado”. *Por ejemplo*, a qué partido votó en las tres últimas elecciones; qué trabajos ha rea-

lizado desde que concluyó sus estudios; qué opinión tenía de los inmigrantes hace diez años y ahora. El problema fundamental de estos estudios retrospectivos estaría en la “memoria” o recuerdo, más probable cuanto mayor sea el lapso de tiempo transcurrido y menor relieve tenga el tema sobre el que se pregunta. Ello repercute en la exactitud de la respuesta y en su omisión.

EJEMPLO DE DISEÑO LONGITUDINAL PANEL

Un ejemplo clásico de este tipo de *diseño* es la investigación que Lazarsfeld, Berelson y Gaudet llevaron a cabo en EE UU, durante las elecciones presidenciales de 1940 (la campaña electoral Wilkie-Roosevelt): *The People's Choice: how the voter makes up his mind in a presidential campaign* (publicado en 1944, en la editorial de la Universidad de Columbia).

El estudio tenía por *objetivo* principal analizar las intenciones de voto de los electores del condado de Erie (Ohio), en las elecciones presidenciales de 1940. Para ello, los autores pasaron tres veces (entre mayo y noviembre de aquel año) un mismo cuestionario a la misma *muestra* de electores que extrajeron (siguiendo procedimientos *aleatorios*) a comienzos de la investigación: 600 personas. La primera vez, antes del nombramiento de los electores; la última, momentos antes de la elección. Mediante este seguimiento de la *muestra* querían comprobar el efecto de la campaña electoral en la intención de voto: las mismas personas son seguidas en el transcurso de la campaña electoral.

El extraer información de las mismas personas en diferentes momentos permite analizar mejor el efecto de la campaña electoral en sus preferencias de voto: si éste es modificado o es consecuente al desarrollo de la campaña electoral. Pero, para evitar cualquier explicación alternativa debida a la aplicación repetida de un mismo cuestionario a las mismas personas en diferentes momentos temporales (lo que puede afectar a las respuestas en mediciones posteriores), en las tres ocasiones en las que se pasó el cuestionario, se administró igualmente a una *muestra* alternativa de electores. Aunque pertenecían a la misma población de estudio, las unidades de esta *muestra* diferían en cada fase de recogida de información. Mediante procedimientos aleatorios en cada ocasión se elegía *ex profeso* una muestra diferente de electores para que actuasen de *grupo de control* para eliminar explicaciones alternativas a la relación causal analizada: la influencia de la campaña electoral en el voto. Muestras de individuos diferentes, pero de un mismo perfil sociodemográfico, que permitían comprobar los efectos de aplicar el mismo cuestionario, a las mismas personas, de forma reiterada.

C) Diseños preexperimentales, cuasiexperimentales y experimentales

Probablemente una de las tipologías más conocidas e influyentes en la investigación social cuantitativa sea la que Donald Campbell y Julian Stanley (1970) publicaran por primera vez en 1963 (*Experimental and Quasi-experimental Designs*

for Research), y con posterioridad revisada por Cook y Campbell (1977, 1979). Una tipología que responde a cinco criterios principales de clasificación:

- Cómo se seleccionan las *unidades* de observación.
- *Número* de observaciones realizadas.
- *Grado de intervención* del investigador en el proceso de investigación.
- *Control de variables* explicativas alternativas a las analizadas (*validez interna*).
- Posibilidad de *generalización de los resultados* de la investigación a otros contextos espaciales y temporales (*validez externa*).

a) *Diseños preexperimentales o correlacionales*: se distinguen por tres características principales:

- *Ausencia de manipulación* de las variables intervinientes en la investigación. El investigador se limita a observar el fenómeno que analiza, sin introducir ninguna modificación o alteración en el mismo.
- Sólo se efectúa una *medición* del fenómeno, aunque incluya distintos aspectos del mismo.
- *Ausencia de control* de posibles fuentes de invalidez de la investigación. Ello le resta poder explicativo. Si bien, el uso de técnicas analíticas multivariantes permite que estos diseños también lleguen al establecimiento de relaciones causales, por medio del *control a posteriori*; no cuando se diseña la investigación (a diferencia de los diseños *experimentales*), sino en la fase de análisis.

El no cumplimiento de los requisitos de la *experimentación* (existencia de al menos un grupo de *control*, *aleatorización* en la formación de los grupos y medición *antes y después* del tratamiento), motiva su consideración de “preexperimentos”. Incluye el diseño de un único grupo con sólo *una medición* (posterior al tratamiento), diseño *pretest-posttest de un único grupo*, y cuando se compara entre un grupo *experimental* y otro de *control*, pero sin ninguna medición previa (lo que dificulta el establecimiento de relaciones causales).

Por ejemplo, la encuesta convencional (no la *encuesta panel*), en la que se produce una única medición de la realidad social.

b) *Diseños experimentales*: se caracterizan por el máximo control de posibles fuentes de invalidación de la investigación mediante:

- *Manipulación experimental*: el investigador manipula a priori (antes de la recogida de información) las variables que pueden influir en la ocurrencia del fenómeno que analiza.
- *Formación de grupos de control* totalmente equivalentes al grupo *experimental* (salvo en las variables cuyos efectos se trata de medir).

- *Aleatorización*: los individuos se asignan al azar al grupo *experimental* y de *control* (para garantizar la equivalencia inicial de los grupos).

Estas actuaciones favorecen el análisis de relaciones causales, al poderse eliminar posibles explicaciones alternativas a las observadas (validez *interna*). Su desventaja principal es la carencia de validez *externa* debido a la alteración de la realidad que se analiza (*manipulación experimental*) y a los casos que se observan (de escaso número y no elegidos al azar entre los integrantes de la población de estudio, como se verá en el capítulo 3).

- c) *Diseños cuasiexperimentales*: a caballo entre los dos anteriores. Se diferencian de los diseños *experimentales* en que rara vez acontecen en el marco de un laboratorio (sino en el contexto de la vida real). También, porque la distribución de las unidades de observación entre el grupo *experimental* y de *control* no se realiza al azar (no pudiendo garantizar la equivalencia inicial de los grupos que se comparan). Esta última limitación puede obviarse mediante el *control a posteriori* (aplicando técnicas analíticas multivariantes). Y, a diferencia de los diseños *preexperimentales*, el investigador no se limita a observar, sino que puede estructurar la situación para facilitar su análisis.

En función de estas características, se distinguen los siguientes diseños *cuasiexperimentales*:

- Diseños donde se *manipula* la situación experimental, pero se carece de grupo de *control*.
- Diseños donde no se *manipula* la situación experimental, pero sí existen grupos de *control* y *experimental* equiparables.
- Diseños con *grupo de control* no equivalente.
- Diseños de *series temporales* (a partir de la información extraída de fuentes estadísticas y documentales, de encuestas periódicas, o mediante otra estrategia de investigación). La variable *dependiente* se mide en distintos períodos de tiempo para analizar su evolución.

Variaciones de estos diseños proceden de los *entornos* donde se extraen los datos (*naturales*, *seminaturales* y *artificiales*) y de los “artefactos” sociales o *rastros* de actividades sociales dejados por sus participantes, siendo fuente de datos principal para los historiadores (de ellos se habla en el capítulo 4, dedicado a la investigación “no reactiva”). Si en las investigaciones en entornos *naturales* la información se recaba mediante observación directa, en los *seminaturales* mediante la respuesta que dan los informantes sobre sus actitudes, opiniones, actividades..., y acerca de otros. Probablemente sea ésta la forma más habitual de investigación en las ciencias sociales. A ellas se suma la realizada en entornos *artificiales*. Éstas no se limitan a los diseños *experimentales* (descritos más pormenorizadamente en el capítulo 3). Se extienden a cualquier investigación donde los datos se recaban en entornos “controla-

dos”, no naturales, sino simulados. Incluyen *simulaciones* y *juegos*, que permiten analizar rasgos de la vida social en situaciones controladas.

- En los *juegos* los participantes pueden manipular algunos factores de su actividad. Difiere de los *experimentos* en que no se enfoca tanto al análisis de la causalidad, sino a replicar algunas situaciones sociales reales y comprender la complejidad de los procesos e interacciones sociales. La teoría de *juegos* analiza situaciones estratégicas en que los agentes toman decisiones en función de sus expectativas sobre lo que van a hacer los demás. Constituye uno de los elementos básicos de la *Teoría sobre la elección racional* (Sánchez-Cuenca, 2009).
- La *simulación* como método de analizar los fenómenos sociales se ha potenciado gracias a los avances en programación e informática (Gullarhorn y Gullarhorn, 1963; Gilbert, 1993; Gilbert y Troitzsch, 2006). Se desarrolla una representación de algún fenómeno social (programable en un ordenador) para comprender algunas características del mundo social o para la predicción (simulando el transcurso del tiempo. Un ejemplo es su uso en la investigación demográfica: usar un modelo que incorpore tasas de mortalidad y fertilidad por edades para predecir cambios demográficos dentro de una década.

EJEMPLO DE SIMULACIÓN

Simulación realizada por José I. García-Valdecasas Medina (2011: “La simulación basada en agentes: una nueva forma de explorar los fenómenos sociales”. *REIS*, 136: 91-110).

Se define la *simulación* basada en agentes como un nuevo método de investigación que permite tratar de manera sencilla la complejidad, la emergencia y la no linealidad típica de muchos fenómenos sociales. El autor lo ilustra a partir de un ejemplo clásico: el *modelo de segregación urbana de Schelling*, que intenta explicar cómo es posible que surjan guetos en ciudades cuyos habitantes no tienen prejuicios étnicos.

El modelo consiste en una cuadrícula de celdas que representa una zona urbana. Cada celda puede estar en tres estados distintos (representados por tres colores): ocupada por un agente de una etnia (color rojo); ocupada por un agente de otra etnia distinta (color verde); y no ocupada por nadie (color negro). No puede haber más de un agente en la misma celda. El algoritmo de la *simulación* empieza distribuyendo los agentes de forma aleatoria entre las celdas que representan la zona urbana. Posteriormente, selecciona una celda al azar. Si la celda está ocupada por un agente, entonces se determina el número de vecinos (los individuos que ocupan las 8 celdas de alrededor) del mismo color: si dicho número es igual o mayor al umbral de tolerancia, entonces el agente está satisfecho y permanece en su misma celda. Por el contrario, si tal número es menor que el umbral de tolerancia, entonces el agente se siente insatisfecho e intenta encontrar una celda que esté ocupada y que tenga un vecindario que le satisfaga. En el caso de que la encuentre, se mueve hacia dicha celda, y si no, permanece en la misma celda. El proceso se repite hasta que todos los individuos estén satisfechos con el vecindario donde viven.

La suposición convencional es que en barrios donde los individuos tienen un umbral de tolerancia mayor del 50% (intolerantes) suelen desarrollar guetos. Por el contrario, en barrios con umbrales de tolerancia menores del 50% (tolerantes) no suelen formar guetos. Sin embargo, al ejecutar el *modelo de Schelling*, se obtuvo que: 1) para umbrales de tolerancia menores del 30% (muy tolerantes) se alcanza el equilibrio rápidamente de forma que todos los individuos están satisfechos con el vecindario donde viven, y no se forman guetos; 2) para umbrales de tolerancia entre el 30 y el 80% se alcanza también el equilibrio y aparecen guetos; 3) para umbrales de tolerancia mayores del 80% (muy intolerantes) no se alcanza el equilibrio y la distribución de individuos es similar a una distribución al azar.

La importancia de este modelo —destacado por el autor— radica en que es posible explicar, a partir de las reglas de conducta de los agentes individuales, resultados sociales que, en principio, no son evidentes o fácilmente predecibles. Pero, como bien reconoce, la *simulación* basada en agentes no es la panacea de la investigación social. Tiene algunos problemas importantes derivados de la verificación y la validación de sus modelos. A veces es difícil saber si los resultados de la *simulación* se deben a causas reales o a defectos en la programación (el problema de la *verificación*): diferentes programas sobre la misma cuestión podrían tener resultados muy distintos. Otras veces es difícil saber si el modelo representa realmente al fenómeno que pretende representar (el problema de la *validación*), pues a menudo no hay suficientes datos empíricos para evaluar la adecuación del modelo con la realidad.

Cualquiera de los *diseños* de investigación puede conllevar la aplicación de una o varias *estrategias* de investigación y *técnicas* de recogida y de análisis de la información. Los datos que se analizan pueden ser:

- *Primarios*: son recabados por el propio investigador (de su objeto de estudio).
- *Secundarios*: han sido recogidos por otro y se utilizan en su forma bruta (matrices de datos, transcripciones de entrevistas...).
- *Terciarios*: datos secundarios que han sido previamente reelaborados o analizados por otros (disponibles en forma de tablas, resúmenes...).

Cada tipo de datos implica un grado diferente de “control” por parte del investigador. “Cuanto más distante esté del proceso de recogida, más difícil le será juzgar la calidad de los datos y garantizar que son apropiados para el proyecto” (Blaikie, 2010: 23). Tendrá que hacerlo explícito en su diseño de investigación, como cualquier problema asociado a sus decisiones y actuaciones para su resolución. Los criterios específicos de evaluación de la “calidad” de los datos *secundarios* (extensibles a los datos *terciarios*), se tratan en el capítulo 4, dedicado a la investigación *no reactiva*. Sobre la calidad de la investigación cuantitativa en general, en el apartado siguiente, con el que concluye la visión panorámica de la metodología *cuantitativa*.

1.5. La calidad de la investigación social

En los últimos años se ha acentuado el interés por la mejora de la “calidad” de la investigación social. Muestra son los artículos publicados en revistas científicas (por ejemplo, el monográfico de la revista *International Journal of Social Research Methodology*, de abril de 2006; pero no es el único). Aunque en cada uno de los capítulos que componen el libro hay un apartado correspondiente a la “calidad” respecto a la temática abordada, en éste se va a ofrecer una visión genérica concerniente al diseño de la investigación.

Sin duda el primer criterio para evaluar un *diseño* de investigación es su adecuación a los *objetivos* principales de la investigación. Si éstos logran cubrirse con la investigación diseñada. En caso negativo, la investigación quedaría desaprobada, ya que el fin para el que se diseñó no logra satisfacerse. Cuando los objetivos sí se cumplen, procede pasar a otros criterios de calidad.

El cumplimiento de los *objetivos* se analiza junto con los otros dos determinantes claves, que inciden en las diferentes fases de la investigación: los *recursos* (materiales, económicos y humanos) y el *tiempo* disponible para formalizar la investigación. Ambos inciden en su buen término y consiguiente consecución de los *objetivos* propuestos. La escasez de recursos y la premura temporal afecta negativamente a la *calidad* de la investigación. Lo mismo no cabe decir a la inversa. La “abundancia” de recursos y tiempo no necesariamente conlleva aumento de la *calidad*. Depende de cómo se hayan administrado en el diseño y la ejecución de la investigación. Dificultades presupuestarias o problemas de tiempo para llevar a cabo la investigación inicialmente diseñada pueden llevar a su modificación e, incluso, poner en peligro el logro de los *objetivos* fijados al inicio de la investigación.

Al cumplimiento de los objetivos se suman otros criterios de evaluación. Los de más amplia aplicación en la metodología cuantitativa corresponden a los inicialmente propuestos por Campbell y Stanley (1970); ampliados por Cook y Campbell (1977) y Reichardt y Cook (1979). Atañen a cuatro criterios de *validez*:

1. *Interna*: el criterio de calidad prioritario en diseños de investigación explicativos, que analizan la “causalidad” de hechos, conductas, opiniones, actitudes u otro fenómeno que trate de explicarse. Para Campbell y Stanley (1970) es el “*sine qua non*” de cualquier indagación empírica o, como dijese Fink (1995: 56), “un diseño ha de ser internamente válido para ser externamente válido y producir hallazgos adecuados”. Hay validez *interna* cuando se han “controlado” explicaciones alternativas a la relación causa-efecto que se analiza.

Por ejemplo, cuando se estudia la relación entre *asistencia a clase* (variable *independiente*) y *calificación* en la asignatura (variable *dependiente*), para poder concluir que entre ambas variables existe relación positiva (a mayor asistencia, mayor calificación), previamente habrá que controlar posibles explicaciones alternativas. Cuantas más variables *perturbadoras* se hayan controlado, mayor validez *interna* tendrá la investigación. Indistintamente de las horas de estudio, del interés por la

asignatura o del cociente de inteligencia, la frecuencia de la asistencia a clase determina la calificación académica.

El *control* puede hacerse *a priori* (antes de la recogida de información), o *a posteriori* (en la fase de análisis). Lo primero es característico de los diseños experimentales (mediante la creación de grupos de control plenamente equivalentes al experimental salvo en el tratamiento aplicado), en tanto que lo segundo, de los *correlacionales* (a través de la aplicación de técnicas de análisis *multivariantes* o análisis cruzados controlando por terceras y cuartas variables). Cuantas más variables perturbadoras se controlen, mayor validez *interna* tendrá la investigación. Aunque el control pleno de cualquier suceso interno o externo a la investigación, que pueda afectar a los resultados, es difícil de lograr. Éste puede también alcanzarse mediante la realización de mediciones reiteradas del mismo fenómeno. *Por ejemplo*, en encuestas *panel*: las personas son las mismas; lo que varía son las circunstancias en las que se encuentran con el paso del tiempo.

2. *Externa*: la posibilidad de “externalizar” o *generalizar* los resultados de la investigación, tanto a la *población* (de la que se ha extraído la *muestra* que se analiza), como a otros contextos *temporales* y *espaciales*. Dependerá de la tenencia de iguales características que las presentes en el contexto espacial y temporal que se analiza. Afecta el número de unidades de la población observada (el tamaño de la *muestra*) y sus características diferenciales respecto de aquellos que no acaban participando en el estudio. Como se verá en el capítulo 7 (sobre el diseño *muestral*) y en el 5 (en el apartado correspondiente a los errores de *no observación* en la *encuesta*), la selección de las unidades muestrales mediante procedimientos aleatorios favorece la “equivalencia” de la *muestra* a la *población* (a la que se quieren hacer extensivos los resultados de la investigación).
3. *De constructo*: una variante de la validez *interna* que se circunscribe a la adecuación de la *medición* de los *conceptos* centrales de la investigación. La operacionalización realizada ha de cubrir las diferentes *dimensiones* del concepto. De ello se hablará en el capítulo 2, donde se hace un análisis más detallado de las diferentes modalidades de validez de *constructo*, una vez detallados los distintos procedimientos de medición. Su aumento se relaciona con la reducción de errores de *especificación* y la operacionalización *múltiple*.
4. *De conclusión estadística*: una variante de la validez *externa* relacionada con el poder, adecuación y fiabilidad de la técnica de *análisis* de datos que se aplique y su adecuación a los objetivos de la investigación. Afecta a la *significatividad* de los análisis estadísticos (las pruebas de hipótesis), a las posibilidades de inferencia de los modelos estadísticos que se obtengan. Incide el tamaño de la *muestra* y su heterogeneidad. A mayor tamaño muestral, mayor probabilidad de que los análisis alcancen significatividad estadística. También afectan los errores típicos de las estimaciones muestrales. Éstos disminuyen a medida que disminuye la varianza de las variables. Por ello se recomienda incrementar el tamaño de la muestra, formar grupos internamente

poco heterogéneos y adecuar la técnica analítica a los objetivos del estudio y a las características de los datos (como se verá en el capítulo 8).

El cuadro 1.4 sintetiza los criterios de *validez* y las recomendaciones convencionales para su consecución.

CUADRO 1.4. Criterios de validez de los diseños de investigación y actuaciones para su mejora

<i>VALIDEZ</i>	<i>ACTUACIONES PARA SU CONSECUCCIÓN</i>
<i>Interna</i>	Control de explicaciones alternativas: <ul style="list-style-type: none"> – Creación de grupos de comparación equivalentes. – Efectuar varias mediciones. – Controlar todo suceso interno y externo a la investigación que pueda afectar a sus resultados.
<i>Externa</i>	Generalización de los resultados del estudio: <ul style="list-style-type: none"> – Incrementar el tamaño de la muestra. – Selección de las unidades de la muestra mediante procedimientos aleatorios. – Abarcar diferentes contextos temporales y espaciales.
<i>De constructo</i>	Adecuación de la medición de los conceptos teóricos: <ul style="list-style-type: none"> – Delimitación clara y precisa de los conceptos. – Operacionalización múltiple. – Emplear varias técnicas de obtención de información.
<i>De conclusión estadística</i>	Significatividad de los modelos estadísticos: <ul style="list-style-type: none"> – Incrementar el tamaño de la muestra. – Formar grupos internamente poco heterogéneos. – Adecuar la técnica estadística a los objetivos del estudio y a las características de los datos.

A la validez se suma la *fiabilidad* o “consistencia” de los datos del estudio: la obtención de iguales resultados aplicando los mismos procedimientos de medición. La *validez* suele ser garantía de *fiabilidad*, pero no a la inversa. *Por ejemplo*, se puede aplicar el mismo *cuestionario* en diferentes ocasiones, obteniéndose los mismos resultados, y éstos ser inválidos porque no miden actitudes, opiniones o comportamientos “reales”. En ello incide la existencia de diferentes errores de observación o medición (descritos en el capítulo 5), incluidos los *efectos reactivos* (dependen de la voluntad y conocimiento de los participantes en el estudio) y *efectos* debidos a la actuación del *entrevistador*. La *fiabilidad* se analiza en el capítulo 2, a cuya lectura se remite.

Lecturas complementarias

- Alvira Martín, F. (1989/2000): “Diseños de investigación social: criterios operativos”, en García Ferrando, M.; Ibáñez, J. y Alvira Martín, F., *El análisis de la realidad social*. Alianza Universidad. Madrid: 85-111.
- Blaikie, N. (2010): *Designing social research*. Polity Press. Cambridge.
- Cresswell, J. W. y Plano Clark, V. L. (2007): *Designing and conducting mixed methods research*. Thousand Oaks, California.
- Hakim, C. (1994): *Research design: strategies and choices in the design of social research*. Routledge. Londres.
- Johnson, B. y Turner, L. A. (2003): “Data collection strategies in mixed methods research”, en Tashakkori, A. y Teddlie, Ch., *Handbook of mixed methods in social & behavioral research*. Sage. Thousand Oaks, CA: 297-319.
- Neuman, W. L. (2007): *Basics of social research: qualitative and quantitative approaches*. Allyn and Bacon. Boston, MA.
- Teddlie, Ch. y Tashakkori, A. (2003): “Major issues and controversies in the use of mixed methods in the social and behavioral sciences”, en Tashakkori, A. y Teddlie, Ch., *Handbook of mixed methods in social & behavioral research*. Sage. Thousand Oaks, CA: 3-50.
- Wallace, W.L. (1976): *La lógica de la ciencia en la sociología*. Alianza. Madrid.

Ejercicios propuestos

1. A partir de una *idea*, formule el *problema* de investigación. Concrete los *objetivos*, el *marco teórico* y las *hipótesis* principales. En su exposición, cite las referencias bibliográficas consultadas.
2. Localice una investigación que articule diferentes metodologías. Describa la función que cada una de ellas cumple en la investigación.
3. Diseñe un estudio *seccional* y *longitudinal* (en sus tres modalidades) para un mismo *problema* de investigación.
4. Escoja una investigación y describa su *diseño*. Destaque qué criterios de *validez* cumple y posibles fuentes de invalidez.
5. Durante varias semanas se observa el absentismo laboral de un grupo de personas que trabajan de noche en un hospital. Con el fin de reducir dicho absentismo, se incrementa el salario del personal del hospital que trabaja en el turno de noche. A la semana siguiente, y durante un mes, se observa el absentismo. ¿A qué *diseño* de investigación corresponde este ejemplo? Especifique las diferentes “amenazas” a su *validez* y actuaciones para su reducción.

2

LA MEDICIÓN DE LA REALIDAD SOCIAL

Las ciencias sociales, al igual que las naturales, precisan de la medición de la “realidad” mediante instrumentos que sean *válidos* y *fiables*. Pero a diferencia de la realidad física, la *social* añade complejidad al proceso de *medición*. Se cuentan con medidas precisas de peso, volumen, altura o temperatura, que pueden aplicarse con iguales parámetros en diferentes contextos sociotemporales. ¿Puede decirse lo mismo de conceptos habituales en la investigación social como discriminación, prejuicio, satisfacción, felicidad, racismo, homofobia...? Como señala Neuman (2000: 157):

“Nosotros también medimos el mundo no físico en la vida diaria, pero normalmente en términos menos exactos. Medimos cuando decimos que un restaurante es excelente, que Pablo realmente es elegante, que Karen tiene una actitud negativa hacia la vida, que Johnson realmente es prejuicioso, o que la película de anoche tenía mucha violencia. Sin embargo, tales juicios diarios como “realmente prejuicioso” o “mucha violencia” son imprecisos, vagos o medidas intuitivas”.

La medición de la temperatura con un termómetro, o del peso mediante una báscula, en cambio adquiere una mayor precisión. El presente capítulo trata de la *medición* y de las fases que comprende. Medición que permite graduar lo que de otra forma sería invisible.

2.1. Fundamentos de la medición

Una sinopsis de aportaciones clave en la fundamentación de la *medición* de la realidad social ha de abarcar contribuciones de ámbitos diferentes. Con riesgo de cometer omisiones relevantes, se destaca:

- La aportación pionera de George A. Lundberg, de 1939 (*Foundations of Sociology*. Macmillan. Nueva York) a la conceptualización del *operacionismo*, a partir de las propuestas de P. W. Bridgman en el campo de la física (expresa en su obra *The Logic of Modern Physics*, también editada por Macmillan en 1927). Para Lundberg la *definición operacional* es el *constructo*; no un proceso para observar un *constructo* después de que se ha definido. Supone que cualquier cambio en la *definición operacional* implica un nuevo *constructo* teórico.

La aportación de Lundberg a la sociología se incluye dentro del *neopositivismo* y la crítica a la Escuela de Chicago. Fue famosa la controversia en torno a la *operacionalización* de conceptos entre Lundberg (1942: “Operational definitions in the social sciences”. *American Journal of Sociology*, 47) y Herbert Blumer (1931: “Science without concepts”. *American Journal of Sociology*, 36), perteneciente a la Escuela de Chicago. Mientras que el primero defendía la operacionalización incluso aunque el *concepto* no estuviese plenamente definido, el segundo disenta y abogaba por demorar la *operacionalización* hasta no haber alcanzado una correcta *conceptualización*.

- La contribución clave de Paul F. Lazarsfeld, en su artículo inicial “Evidence and inference in social research”, de 1958 (*Daedalus*, 87: 99-130), y en publicaciones posteriores (1971, 1972 y 1973). Configura el procedimiento a seguir para ir de la *conceptualización* a la *operacionalización* de conceptos teóricos.
- Hans Zeisel (1957: *Say it with Figures*. Harper and Brothers Publishers. Nueva York) destacó qué aspectos han de considerarse para elaborar índices: exactitud (validez), utilidad teórica y práctica, economía y claridad.
- Stanley S. Stevens (1946: “On the theory of scales of measurement”. *Science*, 103: 677-680), a quien se debe la distinción de los niveles de medición o escalas de medida en nominales, ordinales, de intervalo y de razón. Al igual que la definición de la *medición* como “asignación de números a objetos de acuerdo con determinadas reglas” (1951; *Handbook of Experimental Psychology*. John Wiley. Nueva York, p. 22).

En el ámbito concreto de la *medición de las actitudes*, sobresalen las siguientes propuestas (que se detallan en el apartado 2.4) de acuerdo al siguiente orden cronológico:

- Emory S. Bogardus (1925) “Measuring social distances”. *Journal of Applied Sociology*, 9: 299-308 / (1933) “A social distance scale”. *Sociology and Social Research*, 17: 265-271.
- Louis L. Thurstone (1928) “Attitudes can be measured”. *American Journal of Sociology*, 33: 529-544 / (1931) *The measurement of social attitudes*. University of Chicago Press. Chicago.
- Louis L. Thurstone y E. J. Chave (1929) *The measurement of attitudes*. University of Chicago Press. Chicago.

- Rensis Likert (1932) “A technique for the measurement of attitudes”. *Archives of Psychology*, 140: 44-53.
- Louis Guttman (1944) “A basis for scaling qualitative data”. *American Sociological Review*, 9: 139-150 / (1947) “The Cornell techniques for scales and intensity analysis”. *Educational and Psychological Measurement*, 7: 247-280.
- Charles E. Osgood (1952) “The nature and measurement of meaning”. *Psychological Bulletin*, 49: 197-237.
- Charles E. Osgood y Percy H. Tannenbaum (1955) “The principle of congruence in the prediction of attitude change”. *Psychological Bulletin*, 62: 42-55.
- Charles E. Osgood, George J. Suci y Percy H. Tannenbaum (1957) *The measurement of meaning*. University of Illinois Press. Urban III.

Asimismo cabe destacar el famoso artículo de Richard LaPiere “Attitudes vs. Actions”, de 1934 (*Social Forces*, 13: 230-237), que cuestiona la validez de las respuestas verbales en la medición de las actitudes. Mediante un cuestionario se obtiene una reacción verbal ante una situación enteramente simbólica. Reacción que necesariamente no coincide con la actuación real del encuestado cuando se encuentra ante dicha situación. También resaltan las propuestas de:

- *Operacionalización múltiple* de Donald T. Campbell y Donald W. Fiske (1959: “Convergent and discriminant validation by the multitrait-multimethod matrix”. *Psychological Bulletin*, 56: 81-105), como una estrategia para aumentar la validez de constructo.
- Delimitación de la validez de constructo de Lee J. Cronbach y Paul E. Meehl (1955: “Construct validity in psychological tests”. *Psychological Bulletin*, 52: 281-302), un concepto clave en la teoría psicométrica que se detalla (junto a otros criterios de validez y fiabilidad) en el apartado 2.5.

2.2. El proceso de medición: de los conceptos teóricos a los indicadores empíricos

Antes de “medir” se precisa una clara delimitación del “objeto de interés”. De acuerdo con Becker (1998), éste es el comienzo de la *conceptualización*: ver en qué se está interesado en medir y en qué no. Algunos objetos serán fáciles de medir, por su componente objetivo (*por ejemplo*, la edad, el color de la piel, el nivel de estudios...). Por el contrario los subjetivos revestirán de mayor dificultad (*por ejemplo*, actitudes, valores, percepciones, prejuicios, anomia, cohesión social, amistad...). Pero, ¿qué es un concepto?

- “Símbolos lingüísticos que categorizan a los fenómenos” (Phillips, 1985: 77). Pueden derivarse de reflexiones teóricas, de generalizaciones empíricas (que derivan de investigaciones), o de reflexiones propias sobre la variedad de ex-

perencias de la vida social. Representan *constructos* abstractos, que no directamente observables.

Por ejemplo: los conceptos de *anomia*, *cohesión social*, *amistad*, *discriminación*, *xenofobia*.

- “Una familia de ideas” (Kaplan, 1964: 49). “Categorías para la organización de ideas y observaciones” (Bulmer, 1992: 43).

Por ejemplo: el concepto de *compasión* es un *constructo* (o concepción abstracta de la realidad) creado a partir de la idea de *compasión*, que comparten quienes han utilizado dicho término. No puede observarse porque no existe; se construye.

- Una “abstracción” formada por la generalización de fenómenos similares. Incluso cuando su nombre es una palabra comúnmente utilizada (como *clase social*, *desempleo*, *satisfacción*) se adaptan o crean por los investigadores en consonancia con sus objetivos de investigación. El *constructo*, por el contrario, es un concepto sistemáticamente definido, más formal, para su uso en la teoría científica (Kerlinger, 1986; Hox, 1997).

Por ejemplo: *delincuencia* es un *concepto* que puede definirse haciendo referencia a conductas que cada sociedad cataloga de “delictivas”. La *desviación* sería un *constructo*, cuando se define en el marco de alguna teoría que delimite qué conductas son “desviadas”.

- Los *conceptos* cumplen una función de “síntesis”, de denominación común, que engloba a una serie de observaciones, proporcionándoles un sentido.

Por ejemplo: el concepto de *tristeza* ofrece una denominación común a una serie de manifestaciones diferentes: llanto, desánimo, bajo tono de voz, negativa a hablar o a comer, problemas de sueño...

- Los *conceptos* difieren por su grado de abstracción.

Por ejemplo: el concepto de *poder* connota mayor abstracción que los conceptos de *educación* o *empleo*.

Los *conceptos* constituyen variables “latentes”, al no ser directamente observables. Para poderse medir, han de traducirse en variables “empíricas” (o *indicadores*) que permitan calibrar el grado de ocurrencia del *concepto* que miden. La vinculación de los *conceptos* a *indicadores* se realiza a través del proceso de la *operacionalización* y consiguiente *medición*: la concreción de las variables y de los valores que incluyen. Por el contrario, la *conceptualización* se define como “la formación de

conceptos, que establece el significado de un constructo, elaborando la red nomológica y definiendo importantes subdominios de su significado” (Hox, 1997: 53).

Lazarsfeld (1958, 1971, 1973) concreta el procedimiento de la *conceptualización* a la *operacionalización* (de conceptos teóricos) en cuatro fases:

1. *Representación teórica del concepto*, de forma que queden reflejados sus rasgos definitorios.
2. *Especificación del concepto*, descomponiéndolo en las distintas *dimensiones* o aspectos relevantes que engloba.
3. *Para cada una de las dimensiones se selecciona una serie de indicadores* (o variables empíricas), que “indiquen” la extensión que alcanza la dimensión en los casos analizados.
4. *Síntesis de los indicadores mediante el cálculo de índices*. A cada *indicador* se le asigna un *peso* o valor, de acuerdo con su importancia. A partir de estos valores se confecciona un *índice* (o medida común que agrupe varios indicadores de una misma dimensión numéricamente operacionalizada).

Fiske (1971) también parte de la representación teórica del *concepto*. Defiende que el *núcleo teórico* es la esencia del *concepto*, y que ha de cubrir todos los fenómenos a los que el concepto se refiere. Ello exige:

- Establecer, de manera explícita, lo que el *concepto* incluye y excluye, y las diferencias con otros conceptos relacionados.
- Describir su polo opuesto.

Por ejemplo, si quiere medirse el *concepto* de satisfacción, habría que definir su opuesto: la insatisfacción. Al igual que sucede con los conceptos de pobreza-riqueza o de integración-discriminación.

Hox (1998) reitera la necesidad de:

- *Clarificar la definición constitutiva del constructo*, asemejándose al análisis del significado de *núcleo* de Fiske (1971). La definición ha de ser:
 - *Adecuada*: contener suficientes características que permitan una delimitación satisfactoria de los límites empíricos y referentes del constructo (definición *denotativa*).
 - *Parsimoniosa*: no incluir ninguna característica innecesaria.

Han de reducirse los problemas de *connotación*, de confusión de significados, porque el *constructo* esté asociado a más de un significado o porque dos constructos apunten al mismo significado (y los investigadores no hayan explicitado lo que pretenden medir).

- *Determinar qué referentes empíricos pertenecen al constructo* y cuáles no. Conviene evitar problemas de *denotación*: constructos “vagos”, que no estén definidos con suficiente precisión (a qué objetos o referentes se aplican). Por lo que ha de restringirse su significado, incrementando el número de las características que los definen.
- *Asegurarse de que la etiqueta verbal dada al constructo sea comprendida* de manera inequívoca. Los términos abstractos con frecuencia tienen pocas características distintivas y se refieren a un gran conjunto de referentes empíricos. Los problemas *terminológicos* han de resolverse: elegir palabras que sean claras y precisas, que distingan al constructo de otros próximos.

A la *conceptualización* puede llegarse desde la *teoría*, a través de un proceso *deductivo* (de la teoría a los datos, de lo general a lo concreto), del que se extraen los *constructos* y se seleccionan los *indicadores* (o variables observables). También puede llegarse desde los *datos* empíricos, mediante la operación metodológica de la *inducción* (de lo concreto a los principios generales, de los datos a la teoría), como se vio en el capítulo 1. Procediendo a partir de un proceso *deductivo*, los *conceptos* y las proposiciones que conforman una *teoría* se concretan en variables empíricas o *indicadores*, que posibiliten la contrastación empírica de la teoría. A este proceso se le conoce como *operacionalización de conceptos teóricos*. Comprende las cuatro fases siguientes:

1. *Delimitación de los conceptos* mediante definiciones *teóricas* (o nominales) y *operativas* (u operacionales). Mientras que las primeras son más abstractas, las segundas especifican el significado del concepto y cómo se medirá.

Por ejemplo: el concepto de “absentismo laboral”

- Definición *teórica*: “Ausencia del trabajo durante la jornada laboral”.
- Definición *operativa*: “Número de días al mes que el trabajador falta al trabajo sin justificación”.

En la delimitación del concepto es bueno diferenciar lo que “incluye” y “excluye”, sus diferencias con otros conceptos relacionados, y describir su opuesto. *Por ejemplo:* xenofobia-xenofilia; integración-discriminación.

2. *Descomposición del concepto en dimensiones* (o aspectos a incluir). Las dimensiones estarán en función de los objetivos de la investigación. Es preferible (a efectos operativos) acotar el número de dimensiones a las más relevantes para medir el concepto, que abarcar indiscriminadamente un número amplio de dimensiones.
3. *Buscar indicadores* (para cada dimensión) que midan la propiedad latente del concepto que representan. Los *indicadores* han de adecuarse a los objetivos de la investigación. Se recomienda la elección de varios indicadores para

cada dimensión como estrategia para aumentar la *validez de constructo*. Es de utilidad:

- *Elaborar una lista* que incluya el mayor número posible de *indicadores* para poder abarcar la diversa dimensionalidad del concepto que se mide. Después, en función de la información que los indicadores proporcionen, se procederá a eliminar aquellos que resulten no significativos en la medición del concepto.
- *Acudir a indicadores ya validados* en investigaciones previas. Ello favorece la realización de estudios comparativos, además de ayudar a la selección de aquellos que ya han mostrado cumplir requisitos de *validez* y de *fiabilidad* (apartado 2.5). No obstante, ha de valorarse el posible efecto diferencial debido al contexto sociocultural y temporal en el que se aplique.

EJEMPLOS DE OPERACIONALIZACIÓN DE UN CONCEPTO TEÓRICO: DEFINICIÓN - DIMENSIONES - INDICACIONES

A) Operacionalización del concepto “*vinculación social*” a partir de modelos explicativos de la delincuencia juvenil.

Una de las teorías explicativas de la delincuencia juvenil que ha recabado mayor corroboración empírica es la *Teoría del Control Social* de T. Hirschi (1969: *Causes of delinquency*. University of California Press). Esta teoría se fundamenta en tres *axiomas* (o aseveraciones) principales:

1. Los *controles sociales* son los que mantienen al joven en la conformidad.
2. La *interiorización de la norma* constituye el determinante de la conducta delictiva.
3. Los *lazos* o frenos que mantienen al individuo en la conformidad son:
 - *Afecto* a personas convencionales (padres, profesores, amigos).
 - *Compromiso* con líneas convencionales de acción.
 - *Implicación* en actividades convencionales.
 - *Creencia* en la legitimidad. Es de la creencia en la legitimidad de la norma de lo que depende, en último término la “autorepresión” del individuo y su no implicación en la conducta delictiva.

De estos comienzos axiológicos se deducen *proposiciones* de las que derivan cuatro *hipótesis*:

- Cuanto mayor respeto y admiración sienta el joven hacia las personas con las que se identifica, menor es la probabilidad de que emprenda conductas delictivas.

- Cuanto menos *vinculado* esté el joven en la sociedad convencional (familia, escuela, comunidad), más predispuesto estará a transgredir la norma legal.
- Cuanto más ocupado esté el joven en la realización de actividades convencionales, menor tiempo encontrará para cometer actos delictivos.
- La ausencia de creencias que prohíban la realización de conductas delictivas llevan al joven desvinculado a la delincuencia.

Definición operativa de la “delincuencia juvenil”: “todo tipo de conducta contranormativa (que infringe la norma legal establecida) llevada a cabo por las personas menores de 18 años”.

Una *definición teórica* la proporciona cualquier diccionario al uso: “cualidad de delincuente”, “acción de delinquir”, “conjunto de delitos, ya en general o ya referidos a un país, época o especialidad en ellos” (Diccionario de la Lengua Española: RAE).

De la *hipótesis* [“cuanto menos *vinculado* esté el joven en la sociedad convencional (familia, escuela, comunidad), más predispuesto estará a transgredir la norma legal”] deriva el desglose del *concepto* de “vinculación social” en tres *dimensiones* básicas: familiar, escolar, comunitaria (amigos, vecindad), como muestra el cuadro 2.1.

CUADRO 2.1. Operacionalización del concepto “vinculación social”

CONCEPTO	DIMENSIONES	INDICADORES
Vinculación social	Familiar	<p><i>Características familiares:</i></p> <ul style="list-style-type: none"> – Si vive o no con ambos padres. En caso negativo: motivo de la “ruptura familiar”; edad que tenía cuando se produjo; cómo le afectó. – Número de hermanos, lugar que ocupa (orden de nacimiento). – Estatus familiar: nivel de ingresos, estudios y ocupación de los padres y hermanos.
		<p><i>Afecto hacia padres y hermanos:</i></p> <ul style="list-style-type: none"> – Relación con padres y hermanos (cariño, cordialidad, indiferencia). Grado de comunicación entre ellos. – Tiempo que pasa con ellos y tiempo que desearía pasar. – Identificación con padres y hermanos. – Interés de los padres y hermanos con las actividades que realizan, con sus problemas e inquietudes. – Actividades que comparte con padres y hermanos. – Valores familiares: cumplimiento de las normas. Sanciones que suelen aplicarse cuando no se cumplen (castigos físicos, materiales, psíquicos). – Deseos de satisfacer expectativas familiares. – Discusiones familiares: frecuencia, intensidad, relevancia del tema.

CUADRO 2.1. (Continuación)

CONCEPTO	DIMENSIONES	INDICADORES
Vinculación social	Escolar	<p><i>Actuación escolar:</i></p> <ul style="list-style-type: none"> – Último curso académico realizado. Calificaciones obtenidas. – Si ha repetido o no algún curso académico.
		<p><i>Actitud hacia la escuela:</i></p> <ul style="list-style-type: none"> – Motivación al éxito escolar. Interés y afecto a la escuela. Si le gusta ir al colegio. Número de horas que dedica al estudio. – Participación (voluntaria) en actividades escolares. – Aceptación de la autoridad escolar. Si les preocupa la opinión que de él o ella tengan sus profesores y compañeros. – Percepción de sí mismos como estudiantes: <i>autoestima</i>¹. – Aspiraciones educativas y ocupacionales (compromiso con modelos de trabajo convencionales).
	Comunitaria	<p><i>Amigos:</i></p> <ul style="list-style-type: none"> – Número de amigos. Tipo de relación que tiene con ellos. – Si cometen alguna clase de conducta contranormativa. – Grado de identificación con los amigos. – Tiempo que pasa con ellos. <p><i>Vecindad:</i></p> <ul style="list-style-type: none"> – Relaciones de vecindad. – Características de la vivienda donde reside: condiciones de habitabilidad, hacinamiento, tenencia en propiedad, en alquiler, período de tiempo en última residencia. – Características del barrio: densidad de la población, movilidad residencial, heterogeneidad², ubicación en la ciudad, porcentaje de viviendas en propiedad y en alquiler, servicios comunitarios (escuelas, mercados, red sanitaria, equipamientos deportivos y de ocio, espacios verdes, espacios controlados: patios privados, jardines), vigilancia en las calles, iluminación, tasa de delincuencia en la zona. – Tiempo que pasa en la calle.

1. Hirchi (1969: 113) destaca la importancia de la variable *autoestima* en el proceso causal de la delincuencia: “cuanto más competente crea el individuo que él es, menos probable será que cometa actos delictivos”.

2. Estas tres características del barrio fueron destacadas por la *Teoría de la transmisión cultural* de Shaw y McKay (1969), como características de las *áreas de delincuencia*.

B) Operacionalización del concepto “*calidad de vida*” realizado por M.^a Luisa Setién (1993: *Indicadores sociales de calidad de vida. Un sistema de medición aplicado al País Vasco*. CIS. Madrid. Monografía n.º 133).

La *calidad de vida* es un concepto abstracto, complejo e indirectamente medible, que admite diferentes orientaciones y definiciones. Por lo que procede elaborar un *marco conceptual* que delimite los significados, además de proporcionar un marco coherente para el conjunto de los *indicadores sociales* que se deduzcan.

Como la *calidad de vida* se vincula al “desarrollo”, Setién parte de tres teorizaciones principales: la *Teoría del Desarrollo* de Galtung y Wirak (1979), la *Teoría de las Necesidades* de Maslow (1981) y la *Teoría de las Discrepancias Múltiples* de Michalos (1985). A partir de estas teorías proporciona las definiciones siguientes de cada concepto esencial en la investigación:

- *Desarrollo*: “proceso de realización de las potencialidades humanas mediante la satisfacción de las necesidades, utilizando para ello los medios sociales y contando con un entorno natural limitado”.
- *Necesidades*: “aquello que es menester para la vida humana, sea física, mental, espiritual o social”.
- *Calidad de vida*: “grado en que una sociedad posibilita la satisfacción de las necesidades (materiales y no materiales) de los miembros que la componen”.

De la satisfacción de las necesidades dependerá el grado de calidad con que puede calificarse la vida social en las dicotomías: *peor-mejor*, *bueno-malo*. De esta forma el *marco conceptual* de la investigación orienta la medida de “*calidad de vida*” a la satisfacción de las necesidades. Son las *necesidades* las que proporcionan la base para encarar el sistema de *indicadores sociales*. Se quiere establecer unas áreas de necesidad o componentes de la *calidad de vida* que representen el amplio abanico de necesidades. Para su *operacionalización* se opta por:

- Elaborar amplias listas de *necesidades* (físicas, psicológicas, espirituales, sociales, culturales), siguiendo el criterio de la experiencia.
- Analizar comparativamente los sistemas de *indicadores sociales* establecidos con anterioridad, siguiendo el criterio del “consenso político”.

De la conjunción de ambos criterios resulta la selección de 11 áreas o componentes de *calidad de vida*, que parecen abarcar todos los ámbitos de las *necesidades* humanas:

1. Salud.
2. Renta (aspecto material-bienes).
3. Trabajo.
4. Vivienda.
5. Seguridad.
6. Educación-Formación.
7. Familia.
8. Entorno físico-social (medio ambiente físico y medio ambiente social que enmarca la vida cotidiana-relaciones interpersonales, vecinales y de amistad-integración en grupos, organizaciones, en la comunidad).

9. Ocio-tiempo libre.
10. Religión.
11. Política (participación, libertad, prestigio e identidad comunitaria).

A continuación se determinan las *dimensiones* que comprende cada área que compone la “*calidad de vida*”. La *dimensión* se define como una categoría genérica en la que se divide el área, estando referida a algún aspecto importante de la misma. A partir de las dimensiones se seleccionan los *indicadores* que actúen en su medición. En total se diferenciaron 39 *dimensiones* (algunas de ellas incluso divididas, a su vez, en *subdimensiones*) y 251 *indicadores*, que dan contenido al sistema para la medida de la “*calidad de vida*”. Como ilustración, en el *área* de salud se distinguieron las cuatro *dimensiones* siguientes:

1. Estado de salud de la población y su distribución.
2. Atentados contra la salud.
3. Educación sanitaria.
4. Recursos para la salud.

En la primera *dimensión* (“estado de salud”) quiere medirse la duración de la vida de la población en la comunidad. Para ello se utilizan varios *indicadores*:

1. Esperanza de vida (al nacer y a la edad de 1, 20, 40 y 60 años).
- 2a. Tasa de mortalidad.
- 2b. Tasa de mortalidad específica por edades.
3. Tasa de mortalidad según la causa de muerte.
4. Mortalidad perinatal.

Para mostrar la “salud durante la vida”, se seleccionan los siguientes *indicadores*:

5. Morbilidad hospitalaria.
6. Autovaloración del estado de salud.
7. Incapacidad permanente.
8. Incapacidad transitoria.
9. Padecimientos físico/psíquicos.
10. Distribución del estado de salud.

Para la segunda *dimensión* (“atentados contra la salud”) se distinguen seis *indicadores*:

1. Relación peso/talla de los niños.
2. Consumo de tabaco.
3. Consumo de alcohol.
4. Hábitos de ejercicio físico.
5. Nivel de arraigo de modos de vida atentatorios contra la salud.
6. Conocimiento sobre los atentados contra la salud.

Así se va procediendo en cada área y dimensión, en una investigación que ilustra el desarrollo teórico-metodológico del concepto de “*calidad de vida*”; pero no el metodológico-técnico de recogida y tratamiento de la información.

- C) Operacionalización del concepto “*integración social*” del inmigrante a partir de la *Propuesta para contribuir a la elaboración de un sistema de indicadores comunes de integración*. Ministerio de Trabajo y Asuntos Sociales. Dirección General de Integración de los Inmigrantes (2007).

Propuesta de indicadores del *Proyecto I3* (Indicadores de Integración de Inmigrantes), desarrollado desde julio de 2005 a septiembre de 2006, y cofinanciado por el Programa INTI: “*Integration of Third Country Nationals*” (DG de Justicia, Libertad y Seguridad de la CE). La Dirección General de Integración de los Inmigrantes (Secretaría de Estado de Inmigración y Emigración, del Ministerio de Trabajo y Asuntos Sociales de España) fue la entidad responsable del proyecto. Contó con la participación de centros de seguimiento de la inmigración de administraciones públicas estatales, universidades, ONG, gobiernos locales y regionales de 6 países comunitarios: Alemania, Dinamarca, España, Italia, Países Bajos y Portugal.

El objetivo del proyecto era la construcción de un sistema común de indicadores que mida la integración de la población inmigrante en Europa. Se sustentaba en tres enfoques: 1) El enfoque nacional sobre integración de inmigrantes; 2) La integración vista desde las ciudades (el entorno local como medio principal de integración); 3) La integración vista desde los propios colectivos (la población inmigrante, incluyendo una percepción específica de las mujeres, y la sociedad de acogida).

Definición de “*inmigrante*”: “aquella persona nacional de terceros países” (coincide con la definición funcional utilizada por la Comisión Europea).

Definición de “*integración*”: “un proceso bidireccional basado en los derechos mutuos y las obligaciones correspondientes de los ciudadanos de terceros países en situación legal y de la sociedad de acogida, que permite la plena participación de los inmigrantes” (Comunicación de la Comisión al Consejo, al Parlamento Europeo, al Comité Económico y Social Europeo y al Comité de las Regiones, sobre inmigración, integración y empleo. Bruselas, 3.6.2003).

Ello supone que:

- La sociedad de acogida tiene la responsabilidad de garantizar los derechos formales de los inmigrantes para que puedan participar en la vida económica, social, cultural y civil del país.
- Los inmigrantes deben respetar las normas y los valores fundamentales de la sociedad de acogida y participar activamente en el proceso de integración, sin por ello tener que renunciar a su propia identidad.

Al concebir la “*integración*” como un proceso bidireccional se confiere igual atención a las actitudes de las sociedades receptoras, sus ciudadanos, sus estructuras y organizaciones. La *integración* incluye múltiples facetas de la vida económica, social, cultural y civil, tanto en el ámbito público como en el privado. Los principales ámbitos son:

1. Trabajo, empleo y seguridad social.
2. Educación.
3. Salud.
4. Vivienda.

5. Servicios sociales.
6. Conocimiento básico del idioma, la historia y las instituciones.
7. Adhesión a los valores.
8. Acceso a las instituciones y a los servicios tanto públicos como privados.
9. Competencia intercultural.
10. Inclusión de la inmigración en la formulación y aplicación de otras políticas.
11. Interacción entre inmigrantes y ciudadanos de los Estados miembros.
12. Diálogo intercultural, las redes sociales.
13. Participación ciudadana, participación en la sociedad civil, participación social, la participación en el proceso democrático y en la formulación de las políticas y las medidas de integración (especialmente a nivel local).
14. Actitudes de la población.
15. Medios de comunicación.

Los ámbitos o dimensiones referidas definen la gama de áreas sobre las cuales se está actuando en los diferentes países y se perciben como prioritarias para la *integración*.

Tras haber recogido las aportaciones de las tres “visiones” [1) indicadores nacionales propuestos por 6 países; 2) la opinión de los colectivos (población autónoma e inmigrante); 3) indicadores locales referidos a ámbitos específicos de la política municipal], se define un conjunto de *indicadores* comunes que las integre (considerando la red de actores que intervienen en el proceso de integración). La elección de los *indicadores* se asienta en 5 *criterios*:

1. Un sistema de indicadores operativo no debe trabajar con excesiva información (más aún en el marco europeo), porque se colapsaría y los esfuerzos para la búsqueda de información terminarían por agotar el sistema.
2. Los indicadores comunes han de ser de “resultados”, referidos a los logros alcanzados en el proceso de integración. El seguimiento a nivel europeo de la integración de los inmigrantes debería sustentarse en indicadores “fiables” que permitan la comparación entre países, y no sobre la política y los recursos que cada país destina a ello.
3. La existencia de fuentes de verificación y que éstas sean “fiables” y permanentes para poder establecer un seguimiento en el tiempo. No es útil plantearse indicadores que no sean verificables o cuya fuente (por ejemplo, primarias a través de encuestas) no esté disponible de forma permanente.
4. El sistema de indicadores de *integración* debe permitir la comparación entre países. Ha de basarse en fuentes fiables y permanentes para diferente tipología de información. Ello limita los indicadores comunes seleccionados.
5. Las áreas de empleo, educación y lengua son claves para la *integración* de los inmigrantes (recogidas en los principios básicos comunes de la Agenda Común para la Integración de los Nacionales de Terceros países en la Unión Europea: Comunicación de la Comisión al Parlamento del 01/09/2005). Otras áreas, como vivienda y salud, están recogidas en varios de los informes nacionales. Y las áreas de participación social, ciudadana, contactos sociales, sentido de pertenencia y apertura de los servicios públicos, son de especial importancia para los colectivos de inmigrantes y la población autóctona.

De acuerdo con las condiciones y criterios anteriores, se proponen los siguientes *indicadores* comunes, que se presentan en dos cuadros. El cuadro 2.2 establece, por

ámbitos prioritarios, un conjunto de *indicadores* que son básicos, necesarios, realistas (existencia de fuentes disponibles en todos los países) y de ámbito nacional (refleja información estadística de nivel nacional, no local, ni regional). Son *indicadores* inmediatos en el corto plazo:

CUADRO 2.2. Indicadores básicos de “integración”

ÁMBITO	INDICADORES	OBSERVACIONES
<i>Mercado de trabajo y empleo</i>	<ul style="list-style-type: none"> – Tasa de actividad. – Índice de empleo de los inmigrantes. – Tasa de paro - índice de desempleo de los inmigrantes. – Nivel de renta o salario de los inmigrantes. – Número y tipos de contratos laborales de los inmigrantes. – Sectores de inserción laboral de inmigrantes. – Porcentaje de inmigrantes como trabajador por cuenta propia (autónomo, empresario...). – Accidentes laborales en comparación con los autóctonos. 	<p>Han de interpretarse en términos comparativos con la población autóctona.</p> <p>Desagregados por <i>edad</i> y <i>género</i>; inclusive por grupo de <i>origen</i>.</p> <p>Con referencia específica, en algunos indicadores, a la denominada segunda generación o descendientes.</p>
<i>Educación</i>	<ul style="list-style-type: none"> – Índice de asistencia escolar (de los inscritos, cuántos asisten regularmente a la escuela). – Índice de absentismo. – Índice de fracaso escolar: abandono. – Proporción de éxito (finalización de curso y aprobación) en la educación obligatoria: Primaria y Secundaria. – Proporción de inmigrantes con cualificaciones ocupacionales de enseñanza reglada (18-25 años) o de formación profesional. – Proporción de inmigrantes que acceden a la universidad. – Proporción de inmigrantes que finalizan la universidad. 	<p>Han de interpretarse en términos comparativos con la población autóctona.</p> <p>Desagregados por <i>edad</i> y <i>género</i>; inclusive por grupo de <i>origen</i>.</p> <p>Con referencia específica, en algunos indicadores, a la denominada segunda generación o descendientes.</p>
<i>Lengua</i>	<ul style="list-style-type: none"> – Nivel del conocimiento del idioma del país de acogida del alumnado inmigrante en educación obligatoria. 	<p>Se ha optado por referirlo exclusivamente al ámbito de la escuela, dada la imposibilidad de todos los países de obtener datos fiables sobre adquisición y uso de la lengua de los inmigrantes ajenos al sistema escolar. Aunque países como Alemania, Dinamarca o Países Bajos cuentan con exámenes de lengua para la población inmigrante en general.</p>

CUADRO 2.2. (Continuación)

ÁMBITO	INDICADORES	OBSERVACIONES
<i>Vivienda</i>	<ul style="list-style-type: none"> – Índice de viviendas en propiedad en manos de inmigrantes. – Índice de alquileres de inmigrantes. 	<p>Han de interpretarse en términos comparativos con la población autóctona.</p> <p>Desagregados por <i>edad</i> y <i>género</i>; inclusive por grupo de <i>origen</i>.</p> <p>Con referencia específica, en algunos indicadores, a la denominada segunda generación o descendientes.</p>
<i>Salud</i>	<ul style="list-style-type: none"> – Mortalidad diferencial. – Morbilidad diferencial. – Mortalidad infantil. – Proporción de niños inmigrantes vacunados. 	

Además se apuntan “potenciales indicadores en el medio-largo plazo” (cuadro 2.3). Se consideran igualmente importantes en cuanto factor de integración en la sociedad de acogida, pero están aún sujetos a la detección y construcción de fuentes en el conjunto de los países de la Unión Europea (en algunos países ya se miden):

CUADRO 2.3. Indicadores potenciales de “integración”

ÁMBITO	INDICADORES
<i>Participación social y ciudadana</i>	<ul style="list-style-type: none"> – Número de inmigrantes que participan en asociaciones (asociaciones culturales, deportivas, sociales, vecinales, etc.). – Número de asociaciones de inmigrantes (conformadas por personas inmigrantes y que representan sus intereses). – Términos y condiciones de acceso al derecho de voto. – Inscripción efectiva en el censo electoral. – Participación efectiva en las elecciones. – Grado de sindicación de los trabajadores inmigrantes. – Proporción de inmigrantes que participan en partidos políticos.
<i>Contactos sociales</i>	<ul style="list-style-type: none"> – Matrimonios mixtos (matrimonios con parejas del país donde residen).
<i>Apertura de la sociedad receptora</i>	<ul style="list-style-type: none"> – Trabajadores de la administración pública que se desenvuelven en otra lengua de algún colectivo mayoritario de inmigrantes. – Igualdad de oportunidades para acceder a las organizaciones o instituciones dominantes. – Diversidad en la escuela para currículos y libros de texto adaptados.
<i>Sentido de pertenencia</i>	<ul style="list-style-type: none"> – Sentirse seguros, aceptados y no amenazados en la sociedad de acogida. – Conocimiento de la historia, lengua, cultura y hábitos de la sociedad de acogida.

[.../...]

CUADRO 2.3. (Continuación)

ÁMBITO	INDICADORES
<i>Educación</i>	– Nivel del conocimiento del idioma del país de acogida de la población inmigrante.
<i>Mercado de trabajo y economía</i>	– Proporción de inmigrantes en cargos directivos o mandos intermedios en la empresa privada. – Proporción de trabajadores inmigrantes en las administraciones públicas. – Proporción de inmigrantes con capacidad de ahorro. – Proporción de inmigrantes con capacidad de créditos.

4. La *información* proporcionada por los diferentes *indicadores* puede *sintetizarse en un índice*. Éste ayudará a calibrar el grado al que las unidades de observación (casos analizados) se ajusten al *concepto* medido. Su cálculo precisa que los *indicadores* a agrupar compartan una misma escala de medición que posibilite su agregación. Los *indicadores* se ponderarán (se les asignan “pesos”) en consonancia con su importancia para la medición de la *dimensión* latente que incluye el *concepto*. A tal fin es de gran utilidad la aplicación de técnicas analíticas multivariantes, como el análisis *factorial*.

El *Informe FOESSA 1970* resume cuatro recomendaciones esenciales para la obtención de *coeficientes de ponderación*:

- Han de representar, lo más fidedignamente posible, la variable a ponderar.
- Han de ser sencillos. A ser posible, un número entero y pequeño.
- Cuando la variable tenga significados opuestos (de acuerdo-en desacuerdo; satisfecho-insatisfecho), utilizar signos (+) y (–).
- Atributos iguales han de ponderarse de igual manera para posibilitar la comparación. Por ejemplo:

CUADRO 2.4. Ejemplos de coeficientes de ponderación*

<i>Valoración</i>	<i>Coeficiente ponderación</i>	<i>Valoración</i>	<i>Coeficiente ponderación</i>	<i>Valoración</i>	<i>Coeficiente ponderación</i>
Diariamente	30	Muy de acuerdo	3	Muy de acuerdo	+ 2
Casi todos los días	20	Bastante	2	De acuerdo	+ 1
Semanalmente	4	Poco	1	Ni de acuerdo ni en desacuerdo	0
Varias veces al mes	2	Nada	0	En desacuerdo	– 1
Casi nunca	1			Muy en desacuerdo	– 2
Nunca	0				

* Cada coeficiente de *ponderación* se multiplicará por las frecuencias de cada ítem, en cada valoración. Para simplificar y estandarizar ítems diferentes se emplean frecuencias relativas porcentuales (calculadas perpendicularmente a la dirección de la comparación: normalmente en sentido vertical o por columna, como se verá en el capítulo 8).

EJEMPLO DE CÁLCULO DE ÍNDICES

- A) Operacionalización del concepto de “desarrollo social” realizado por J. I. Cano, P. Díaz, A. Sánchez y M. S. Valles (1993: *El desarrollo social de los municipios zamoranos*. Instituto de Estudios Zamoranos “Florián de Ocampo”. Zamora).

El objetivo de la investigación era conocer cómo se estructura la provincia de Zamora económica, social y espacialmente. Primero se analiza el nivel de *desarrollo social* de los municipios de Zamora para, a continuación, establecer una *tipología* de municipios.

En la *operacionalización* del concepto de “desarrollo social” el equipo investigador destaca tres *dimensiones*:

1. *Demográfica*: el perfil demográfico de la población (que refleja otra serie de fenómenos que exceden a la propia demografía).
2. *De accesibilidad y comunicación*: el componente espacial, de especial trascendencia en pequeños núcleos de población. Es el transporte, más que el nivel de equipamientos del propio enclave, lo que más va a condicionar el desarrollo social. Quiere medirse la disponibilidad y el acceso a bienes y servicios.
3. *Socioeconómica, de bienestar y consumo*: analizar el dinamismo económico y social de los municipios zamoranos.

Para la *medición* de cada una de estas tres *dimensiones*, se buscan los *indicadores* que muestren ser más adecuados (dentro de los límites de las fuentes de información disponibles). Las *variables demográficas* incluían *indicadores* de recuento correspondientes a los dos últimos momentos censales (1970 y 1981), padronales (1976 y 1986), *índices* que reflejan la evolución de la población en esos intervalos, e *indicadores* de flujos (*Movimiento natural de la población*) referidos a dichos períodos de tiempo. Se partió de las *variables* siguientes:

- Tasa de jóvenes de los años 1981 y 1986
- Tasa de viejos de los años 1981 y 1986
- Tasa de dependencia de los años 1981 y 1986
- Tasa de masculinidad del año 1986
- Tasa bruta de natalidad de los años 1976-1980 y 1981-1985
- Tasa bruta de mortalidad de los mismos años
- Saldo vegetativo de los años 1976-1980 y 1981-1985
- Saldo migratorio de los mismos años
- Incremento poblacional interpadronal (1975-1986)
- Incremento poblacional intercensal (1970-1981)

A cada *indicador* acompaña su definición, concretando: objetivos, objeciones, representación espacial y fuentes. Después se realiza un análisis de cada uno de los *indicadores* por separado para, a continuación, aplicar la técnica estadística de análisis *factorial* de *componentes principales*. Como se verá en el capítulo 8, ésta permite medir el peso e importancia de cada *indicador* en la *dimensión* considerada, además de la ob-

tención de uno o varios *índices* (o *componentes*) por cada una de las tres *dimensiones*. Ocho *indicadores* mostraron significatividad estadística:

- Tasa de jóvenes de 1986 (TASJO86)
- Tasa de viejos de 1986 (TASVIE86)
- Incremento interpadronal 1975-1986 (INCRPAD)
- Incremento intercensal 1970-1981 (INCRcen)
- Tasa bruta de natalidad de 1981 (TBN81)
- Tasa bruta de mortalidad de 1981 (TBM81)
- Saldo vegetativo medio del período 1981-1986 (SALVEG81)
- Saldo migratorio medio del período 1981-1986 (SALMIG81)

El análisis de *componentes principales* realizado con estas variables define dos *factores* con las saturaciones (“*factor loadings*”) que muestra el cuadro 2.5. Las saturaciones expresan el peso de cada variable en cada *componente*. El signo indica si el *indicador* es directa (signo positivo) o inversamente (signo negativo) proporcional a la puntuación del *componente* (o dimensión latente). La puntuación del *componente* se obtiene multiplicando cada saturación por su indicador correspondiente estandarizado (restándole su *media* y dividiéndolo entre su *desviación típica* para que todos se hallen en la misma unidad de medida) y sumando todos estos términos. En este caso, el primer *componente* explica el 50,8% de la varianza total y el segundo el 16,9%. Al primer *componente* “Factor de Pujanza Demográfica”, por el elevado porcentaje de varianza explicada, y la fuerte saturación que presentan variables cruciales para el crecimiento demográfico (saldo vegetativo e incrementos poblacionales). El segundo *componente* responde más al “saldo migratorio” y, en menor medida, a la tasa de mortalidad y al incremento interpadronal.

CUADRO 2.5. Composición de los componentes

	<i>Componente 1</i>	<i>Componente 2</i>
TASJO86	.84051	.05365
TASVIE86	–.75027	.07639
INCRPAD	.76192	.47921
INCRcen	.63833	.01963
TBN81	.73888	–.00758
TBM81	–.62946	.54007
SALVEG81	.86297	–.36133
SALMIG81	.35076	.83228

Igual procedimiento se sigue con las otras dos *dimensiones* que se diferencian en el *concepto* de “desarrollo social”. A partir de toda la información analizada, se elabora un *índice* de orden superior, en el que se sintetizan los *indicadores* que han obtenido un mayor poder analítico en las tres *dimensiones* (un *índice* sintético de “desarrollo social”). El cuadro 2.6 muestra los *indicadores* e *índices* obtenidos:

CUADRO 2.6. Operacionalización del concepto de “desarrollo social”

CONCEPTO	DIMENSIONES	INDICADORES	ÍNDICES	ÍNDICE GENERAL
<i>Desarrollo social</i>	<i>Demográfica</i>	Tasa de jóvenes Tasa de viejos Incremento interpadronal Incremento intercensal Tasa bruta de natalidad	<i>Pujanza demográfica</i>	<i>Índice de desarrollo social</i>
		Tasa bruta de mortalidad Saldo vegetativo Saldo migratorio	<i>Recesión demográfica</i>	
	<i>De accesibilidad</i>	Distancia a la capital Distancia al núcleo >5.000 h. N.º de autobuses y trenes N.º de vehículos privados Distancia centros educativos	<i>Comunicación / accesibilidad</i>	
	<i>Socioeconómica</i>	Tierras cultivadas Caballos potencia tractores por hectárea cultivada % Tierras en barbecho	<i>Calidad de la tierra</i>	
		Tasa de paro Consumo electricidad de alta tensión por habitante % Población ocupada en sector secundario	<i>Industrialización</i>	
		N.º medio de teléfonos por habitante Consumo de electricidad de baja tensión por hab. % Población con estudios de secundaria	<i>Bienestar y terciarización</i>	

Después de elaborar medidas del “desarrollo social” de forma sectorial, se compone un *índice* general de síntesis del grado de desarrollo. Para ello se realiza un nuevo análisis de *componentes principales* con los *índices* (o componentes) que resultan de los análisis anteriores. Este *índice* (que explica el 56,7% de la varianza total) lo integran cuatro componentes con los siguientes coeficientes factoriales: *pujanza demográfica* (FACDEM: 0.82034), *comunicaciones* (FACOM: 0.78623), *bienestar y servicios* (NFACSERV: 0.78765), *productividad, rendimiento y mecanización de la agricultura*, abreviado con la expresión “calidad de la tierra” (FACALTER: 0.59841). Que éste último correlacione menos con el resto no es de extrañar, debido a que es el único que no se aplica a todos

los municipios por igual (sobre todo a aquellos en los que la agricultura tiene un peso importante). Pero, a diferencia de lo realizado en la creación de los *factores* originales (o *índices*), en la confección del *índice general* los autores no se limitan a estandarizar los componentes y multiplicarlos por sus saturaciones. Multiplican cada componente por una *ponderación* elegida, según la importancia teórica que conceden a cada uno de los *factores*. Las ponderaciones utilizadas son las siguientes: FACDEM ($\times 6$), FACOM ($\times 3$), NFACSERV ($\times 3$), FACALTER ($\times 2$).

Por último, se introduce una lista con las puntuaciones concretas de cada uno de los municipios de Zamora en este *Índice de desarrollo social*. El objetivo de la investigación era conocer la distancia de cada municipio zamorano en términos de “*desarrollo social*”.

B) Ilustración del *cálculo de índices en tablas* tomado de Miguel S. Valles y M.^a Ángeles Cea D’Ancona (1995: “Persona y Sociedad en la Vejez”. En Amando de Miguel, *La Sociedad Española, 1994-1995*. Ed. Complutense, p. 810).

CUADRO 2.7.

Pobl. 65 años y más con hijos (% horizontal)	¿Con qué frecuencia habla por teléfono?					
	A diario	Varias veces a la semana	Una vez a la semana	Una vez al mes	Una vez en meses	Índice* (días/mes)
Varones	40	27	20	7	6	16,11
Mujeres	40	27	20	8	5	16,12
– 75 años	43	27	18	7	5	16,93
+ 75 años	35	26	24	8	7	14,66
Total	40	27	20	7	6	16,11

* El *índice* se ha calculado ponderando cada porcentaje por el coeficiente correspondiente a los días del mes (30, 12, 4, 1, 0) y se ha dividido entre el sumatorio de porcentajes.

Por ejemplo, el *índice* correspondiente a los varones sería:

$$I = \frac{(40 \times 30) + (27 \times 12) + (20 \times 4) + (7 \times 1) + (6 \times 0)}{100} = 16,11$$

El *índice* correspondiente a las personas de 75 años y más:

$$I = \frac{(35 \times 30) + (26 \times 12) + (24 \times 4) + (8 \times 1) + (7 \times 0)}{100} = 14,66$$

C) *Índice de xenofobia y racismo* desarrollado por Juan Díez Nicolás desde el Informe de 1992 sobre “Actitudes de los españoles hacia los inmigrantes” de *La Realidad Social en España* (Centro de Investigaciones sobre la Realidad Social: CIRES) hasta 2007, y revisado en 2009 (“Construcción de un índice de xenofobia-racismo”. *Revista de Trabajo e Inmigración. Migraciones Internacionales*, n.º 80: 21-38).

Las preguntas de la encuesta utilizadas para elaborar el *índice de xenofobia* se refieren a actitudes básicas hacia la inmigración, que se miden con cinco frases de acuerdo-desacuerdo:

1. Solo se debería admitir a trabajadores de otros países cuando no haya españoles para cubrir esos puestos de trabajo.
2. Bastante difícil es la situación económica de los españoles como para además tener que destinar dinero a ayudar a los inmigrantes.
3. Se diga lo que se diga, a todos nos molestaría que nuestros hijos tuvieran compañeros de otras razas en sus escuelas.
4. La inmigración extranjera acabará provocando que España pierda su identidad.
5. Los ciudadanos de cualquier país deberían tener derecho a establecerse en cualquier otro país, sin ningún tipo de limitaciones.

A quienes respondieron estar “muy de acuerdo” o “de acuerdo” con las cuatro primeras frases o “muy en desacuerdo” o “en desacuerdo” con la quinta frase, se les asignó 1 punto. Del mismo modo, se asignó 1 punto a los entrevistados que prohibirían a su hija *tener relaciones con* un hombre de cada uno de los siguientes grupos sociales: gitanos, norteafricanos, africanos de raza negra, sudamericanos y europeos del este. Finalmente, se asignó también 1 punto a quienes juzgaron molesta la *convivencia en el vecindario* con sudamericanos, gitanos, africanos de raza negra y marroquíes. El *índice de xenofobia* así elaborado variaba entre 0 (nada xenófobo) y 14 puntos (muy xenófobo), al ser 14 los ítems que en él participaban.

En su revisión, la aplicación del análisis de *componentes principales* lleva a la obtención de un *índice* (“más económico y sencillo”) con sólo 6 ítems, dos por cada una de las tres dimensiones que resultan del análisis. Se refieren al rechazo o exclusión social de los inmigrantes en el ámbito familiar, de vecindad, laboral y de integración social. En concreto: la prohibición de enamorarse de un norteafricano y de un africano negro, la consideración como vecinos molestos a los africanos de raza negra y a los marroquíes, y el acuerdo con que “sólo se debería admitir a trabajadores de otros países cuando no haya españoles para cubrir esos puestos de trabajo” y con que “bastante difícil es la situación económica de los españoles como para además tener que destinar dinero a ayudar a los inmigrantes”.

En conclusión, mediante la *conceptualización* se clarifican las ideas o constructos teóricos; la *operacionalización* vincula las definiciones conceptuales a indicadores (o variables empíricas); y la *medición* culmina el proceso mediante la atribución de símbolos matemáticos. De acuerdo con Stevens (1951: 22), la *medición* consiste

en “asignar números a objetos de acuerdo con determinadas reglas”. Si bien, la asignación de números ha de hacerse de modo que los números “se correspondan con diferentes grados de *cualidad* –o propiedad– de algún objeto o evento” (Duncan, 1984: 126). Para Carmines y Zeller (1979: 10), la *medición* se centra “en la relación crucial entre el indicador fundamentado teóricamente –eso es, la respuesta observable– y el concepto no observable, latente”. El problema está en cómo concretar esa correspondencia entre lo conceptual y lo empírico para que la *medición* cumpla requisitos de *validez* y *fiabilidad*. Ambos criterios de calidad de la medición se abordan en el apartado 2.5. A continuación se concreta la *medición* de variables.

2.3. Tipologías de variables

La *variable* es “una idea central en la investigación cuantitativa. Simplemente definida, una variable es un concepto que varía” (Neuman, 2000: 126-127). Por *variable* comúnmente se entiende cualquier cualidad o característica de un objeto (o evento) que contenga, al menos, dos atributos (categorías o valores), en los que pueda clasificarse un objeto o evento determinado. Los *atributos* son las diferentes *categorías* o *valores* que componen la variable. En función de ellos se clasifica a los objetos en un grupo u otro. *Variables* como la *edad* (años cumplidos), *distancia* (metros, kilómetros, millas...), *ingresos* (en euros, dólares, francos...), toman *valores* (numéricos). En cambio, las variables *sexo* (varón, mujer), *estado civil* (soltero, casado, en pareja, viudo, separado, divorciado), *satisfacción conyugal* (bastante satisfecho, satisfecho, ni satisfecho ni insatisfecho, insatisfecho, bastante insatisfecho), adoptan *categorías*.

La *medición* de una *variable* consiste, precisamente, en el proceso de asignar *valores* o *categorías* a las diferentes características que conforman el objeto de estudio. Su correcta realización requiere el cumplimiento de tres requisitos básicos:

1. *Exhaustividad*: ha de incluir toda la variedad de *atributos* (categorías o valores) posible para que ningún objeto o sujeto quede sin poder clasificarse. Por ello se recomienda incluir la categoría “otros”, cuando se dude que se hayan cubierto las diferentes opciones de respuesta.
2. *Exclusividad*: los *atributos* han de ser mutuamente excluyentes para que cada observación pueda clasificarse en una sola opción. *Por ejemplo*:
 - Incumplimiento cuando la variable *tamaño de hábitat* se mide en: menos de 2.000 habitantes (1); de 2.000 a 5.000 (2); de 5.000 a 10.000 (3); de 10.000 a 50.000 (4); de 50.000 a 100.000 (5); 100.000 y más (6).
 - Cumplimiento si cambia a: menos de 2.000 habitantes (1); de 2.000 a 5.000 (2); de 5.001 a 10.000 (3); de 10.001 a 50.000 (4); de 50.001 a 100.000 (5); Más de 100.000 (6).

3. *Precisión*: hacer el mayor número de distinciones posibles para lograr una información más precisa. Tiempo habrá para agrupar las distintas *categorías* o *valores* de las variables; generalmente, después de haberse recabado la información (en la fase de análisis), y en función de la frecuencia que presente cada *atributo* de la variable. Lo inverso, en cambio, no es posible: desglosar los *atributos* después de la recogida de la información.

Las *variables* pueden ser clasificadas de acuerdo a cuatro criterios principales: el nivel de medición, la escala de medición aplicada, su función en la investigación y su nivel de abstracción. El cuadro 2.8 sintetiza tipologías de variables conforme a estos criterios de clasificación.

CUADRO 2.8. Tipología de variables según criterios de clasificación

CRITERIO DE CLASIFICACIÓN	TIPO DE VARIABLE		EJEMPLOS
<i>Nivel de medición</i>	<i>No métrica o cualitativa</i>	<i>Nominal</i> Expresa cualidad del objeto o sujeto, sin haber relación de orden entre sus categorías (restriniendo las opciones de análisis estadístico).	Color del pelo, asignatura, estado civil, nacionalidad, partido político. <i>Sexo</i> : Varón 1 Mujer 2 (Los números se asignan de manera arbitraria y no denotan cantidad; sólo pertenencia a la categoría)
		<i>Ordinal</i> Expresa cualidad, pero sus categorías están ordenadas (en el sentido de “mayor que” o “menor que”) – lo que amplía las posibilidades de análisis.	Grado de satisfacción, acuerdo, práctica religiosa, nivel de estudios. <i>Clase social</i> : Alta 5 Media – alta 4 Media – media 3 Media – baja 2 Baja 1 (Ha de haber correspondencia del número con la cualidad que denota)

CUADRO 2.8. (Continuación)

CRITERIO DE CLASIFICACIÓN	TIPO DE VARIABLE		EJEMPLOS	
Nivel de medición	Métrica o cuantitativa	De intervalo	Altura:	
		Expresa cantidad. Puede cuantificarse la distancia exacta que separa cada valor de la variable (amplias opciones analíticas).	Menos de 50 cm 50 – 75 76 – 100 101 – 125 126 – 150 151 – 175 176 – 200 Más de 200	1 2 3 4 5 6 7 8
			(Preferible medir el valor exacto de la variable que su agrupación en intervalos)	
		De razón o proporción	Ingresos:	
		Añade poder establecer el cero absoluto (permitiendo el cálculo de proporciones y de cualquier operación aritmética). Es el nivel de medición más elevado.	Menos de 600 € 600 – 1.000 1.001 – 2.000 2.001 – 3.000 3.001 – 4.000 Más de 4.000	1 2 3 4 5 6
			(Toda variable de intervalo en la que pueda haber cero unidades)	
Escala de medición	Continua	Incluye una escala ininterrumpida de valores (entre dos valores hay opciones intermedias)	Edad: _____ 14 15 16 17... (Entre 14 y 15 años hay infinitos valores: 14 años, 6 meses, días...)	
	Discreta	La escala de medición se interrumpe (entre dos atributos de la variable no caben opciones intermedias)	Toda variable cualitativa (nominal u ordinal) y las cuantitativas que no puedan fraccionarse, como <i>número de miembros</i> de una familia (no puede haber 3 hijos y medio) N.º libros comprados (discreta), pero n.º libros leídos (continua)	

CUADRO 2.8. (Continuación)

CRITERIO DE CLASIFICACIÓN	TIPO DE VARIABLE		EJEMPLOS
<i>Nivel de abstracción</i>	<i>General</i>	No puede observarse directamente, al ser genérica y abstracta. Precisa operacionalizarse	<i>Clase social</i> (su conocimiento preciso exige la confluencia de diferentes indicadores para abarcar su dimensionalidad)
	<i>Intermedia</i>	Expresa alguna <i>dimensión</i> o aspecto que comprende una variable genérica	<i>Ámbitos educativo, laboral y económico</i> (dimensiones comunes en la medición de la clase social)
	<i>Empírica</i>	Directamente medible. <i>Indicador</i> de la dimensión diferenciada en el concepto a medir	<i>Curso académico terminado</i> (dimensión educativa), <i>ocupación principal</i> (dimensión laboral), <i>ingresos</i> (dimensión económica)
<i>Función en la investigación</i>	<i>Dependiente (Y)</i>	Sus atributos “dependen” de los valores de la variable independiente. Su variabilidad es lo que desea analizarse: variable efecto/resultado	<i>Calificación académica (Y)</i>
	<i>Independiente (X)</i>	Sus atributos predicen el valor que adoptará la variable dependiente (por lo que también se llaman variables explicativas o predictoras)	Horas de estudio (X_1) Asistencia a clase (X_2) Coeficiente de inteligencia (X_3) Motivación (X_4) Sexo (X_5)
	<i>Perturbadora</i>	<i>De control</i> Media en la relación entre la variable independiente y la dependiente Su efecto se controla antes o después de la recogida de datos	En la relación de la variable <i>horas de estudio</i> con <i>calificación</i> pueden mediar las otras cuatro variables independientes y cualquier otra que incida en una distinta calificación en personas que han estudiado las mismas horas
		<i>Aleatoria</i> De efecto no previsto, que se reduce cuando la muestra se elige al azar	<i>Ponerse enfermo, tener una mala noticia</i> o cualquier imprevisto que afecte a la calificación del examen, pero en casos esporádicos (no introduciendo sesgos relevantes en los resultados del estudio)

2.4. Escalas de medición de actitudes

La especificidad y complejidad de la medición de las actitudes propició la aparición de diversas propuestas *escalares*. Como los *índices*, las *escalas* “proporcionan al investigador más información sobre las variables y hace posible asegurar la calidad de la medición” (Neuman, 2000: 175). Ambos ayudan a sintetizar la información, conectándose a requisitos de validez y fiabilidad. Si el *índice* es una medida sintética (una puntuación) que combina varios indicadores de un constructo, la *escala* dispone las respuestas u observaciones en un continuo, permitiendo captar la intensidad, la dirección y el nivel del constructo que se mide. Puede utilizar un único o varios indicadores, aunque lo habitual es que agrupe múltiples indicadores y en un nivel de medición ordinal. A continuación se detallan las cinco escalas más populares.

2.4.1. Escala de distancia social de Bogardus

Propuesta por Emory S. Bogardus en 1925 (“Measuring social distance”. *Journal of Applied Sociology*, 9: 299-308) y matizada en 1933 (“A social distance scale”. *Sociology and Social Research*, 17: 265-274), para medir la voluntad de miembros de grupos étnicos diferentes de relacionarse con otros (“distancia social”); aunque puede utilizarse para medir lo próximo o distante que una persona se siente de cualquier otro grupo de población (de diferente etnia, cultura, religión, delincuentes, homosexuales, enfermos de sida...).

Una serie de proposiciones se ponen en orden decreciente, de mayor a menor deseo de interrelación con un grupo de población diferencial, expuesto a estigmas y prejuicios sociales. Se distingue un *grado de intimidad máximo* (matrimonio, noviazgo, amistad), *intermedio* (tenerle como compañero de trabajo, vecino) o *mínimo* (compartir el mismo autobús, su presencia en bares, restaurantes, discotecas) e, inclusive, su *exclusión* en toda modalidad de contacto. La propuesta de Bogardus presenta la siguiente graduación decreciente entre las que elegir la “aceptación” de cada grupo. La puntuación “1.0” indica “ninguna distancia social”, mientras que “7.0”, distancia máxima:

- Me casaría con él (puntuación 1.0)
- Le tendría como amigo (2.0)
- Como vecino en la misma calle (3.0)
- Como compañero de trabajo (4.0)
- Como ciudadano de mi país (5.0)
- Sólo como visitante de mi país (6.0)
- Le excluiría de mi país (7.0)

Es una escala acumulativa, al igual que la de Guttman: el acuerdo con un ítem implica el acuerdo con los ítems precedentes. A partir de las respuestas se puede

elaborar un *índice de distancia social*, en consonancia con los pesos asignados en cada punto de la escala. En su contra, el ser una escala demasiado simple. La *distancia social* puede conceptualizarse desde otros parámetros como la frecuencia de la interacción entre grupos diferentes o las distinciones normativas en una sociedad sobre quién debería ser considerado un “insider” o “outsider” (Karakayali, 2009).

EJEMPLO DE ESCALA DE DISTANCIA SOCIAL DE BOGARDUS

A) En el Informe FOESSA de 1970 (Informe Sociológico sobre la Situación en España), dirigido por Amando de Miguel, se hizo uso de la *escala de distancia social* de Bogardus, para medir la “distancia” hacia personas de: otra región, que han estado durante más de un año en un centro psiquiátrico, pobres y quienes han estado en la cárcel por un delito común. El formato de la pregunta es el mismo en cada supuesto. Lo que varía es el grupo social sobre el que se mide la actitud favorable o desfavorable de los españoles.

La pregunta de *distancia social* hacia las personas de otras regiones (incluida en el cuestionario llamado “Extra-A-Regionalismo”) es como figura a continuación:

P122. SI TUVIESE USTED QUE RELACIONARSE CON LA PERSONA QUE NO ES DE ESTA REGIÓN, ¿CUÁLES DE ESTAS ACTITUDES ADOPTARÍA USTED? (*Enseñar tarjeta*). (*Respuesta única: la primera*).

- | | |
|--|---|
| – La aceptaría en todas las ocasiones e incluso no le importaría casarse con ella | 1 |
| – La aceptaría en todas las ocasiones, excepto como yerno o nuera | 2 |
| – La aceptaría en casi todas las ocasiones como amiga y no tendría inconveniente en invitarla a comer a mi casa (pero no como yerno o nuera) | 3 |
| – La aceptaría como ciudadano y también como vecina, inquilina, compañera de trabajo, etc. | 4 |
| – Sólo la aceptaría como una ciudadana más | 5 |
| – No la aceptaría en ninguna de estas situaciones | 6 |

En el cuestionario Extra F-Psicología se mide la *distancia* hacia un expaciente de un centro psiquiátrico, un expresidiario y un pobre, aplicando la misma escala anterior.

En el *Apéndice Metodológico* del Informe se detalla el procedimiento seguido para obtener un *índice de distancia social*. Los porcentajes de respuesta en cada proposición (o categoría de respuesta) se multiplican por el *coeficiente de ponderación* asignado correspondiente: 5, 4, 3, 2, 1 y 0. El resultado total se divide entre 500 (al haber quedado excluida la sexta categoría por estar multiplicada por 0):

$$\text{Índice de distancia social} = \frac{A\% \times 5 + B\% \times 4 + C\% \times 3 + D\% \times 2 + E\% \times 1}{500}$$

El índice oscila entre un máximo de “+1” (aceptación máxima) y un mínimo de “0” (aceptación mínima), con un punto medio sin significación estadística.

B) Encuestas de *Actitudes hacia la inmigración*.

En las encuestas realizadas por el Centro de Investigaciones Sociológicas (CIS) desde 1990 hasta el barómetro de noviembre de 2005, la *distancia social* hacia personas de diferente procedencia nacional se mide con varias preguntas que, aunque comparten similares supuestos con la escala de Bogardus, presentan un formato diferente, como preguntas separadas que solicitan grado de acuerdo. Las preguntas son las siguientes:

- ¿A Ud. le preocuparía mucho, bastante, poco o nada que un hijo o una hija suya (si no los tiene, en caso de que los tuviera) se casara con un ciudadano de...?
- ¿Hasta qué punto: mucho, bastante, poco o nada, le importaría a Ud. que sus hijos (si no los tiene, en caso de que los tuviera) compartieran en el colegio la misma clase con niños de familias de inmigrantes extranjeras?
- ¿Y hasta qué punto: mucho, bastante, poco o nada le importaría a Ud. tener como vecinos a una familia de ciudadanos de...?
- ¿Y hasta qué punto: mucho, bastante, poco o nada, le importaría a Ud. tener como compañero de trabajo a ciudadanos de...?

Exceptuando la segunda pregunta, en el resto se distingue entre ciudadanos de diferentes países: Portugal, algún otro país de la Unión Europea, Estados Unidos, Europa del Este (Polonia, Hungría, etc.), Latinoamérica, Marruecos u otro país norteafricano, un país del África Negra. Con posterioridad, en las encuestas que el CIS realiza para OBERAXE (Observatorio Español de Racismo y Xenofobia), desde 2007, la *distancia social* se mide mediante un formato de pregunta más próximo a la *escala de Bogardus*. El encuestado ha de responder si “acepta o aceptaría, trataría de evitar o rechazaría” diferentes tipos de relación con inmigrantes en general. No se diferencia por procedencia nacional, distintividad étnica, cultural o religiosa:

P. 28. Ahora le voy a leer una serie de posibles relaciones entre Ud. y personas inmigrantes. Por favor, dígame en cada caso, si acepta o aceptaría ese tipo de relación, trataría de evitarla o la rechazaría.

	Trataría de			Depende (No leer)	NS	NC
	Aceptaría	evitarla	Rechazaría			
– Vivir en el mismo barrio en el que viven muchos inmigrantes	1	2	3	4	8	9
– Vivir en el mismo bloque en el que viven inmigrantes	1	2	3	4	8	9
– Alquilar un piso a inmigrantes	1	2	3	4	8	9
– Trabajar/estudiar con inmigrantes ..	1	2	3	4	8	9
– Que un inmigrante sea su jefe en el trabajo	1	2	3	4	8	9
– Que su hijo/a lleve amigos inmigrantes a casa	1	2	3	4	8	9

– Que su hijo/a se case con una persona inmigrante	1	2	3	4	8	9
– Llevar a su hijo/a a un colegio donde haya muchos niños hijos de inmigrantes	1	2	3	4	8	9

C) Variación de la *escala de Bogardus* aplicada por Dolores Morera y colaboradores (2004: “La percepción de semejanza intergrupal y la identificación con el endogrupo: ¿Incrementa o disminuye el prejuicio?”. *Psicothema*, vol. 16, n.º 1: 70-75) para analizar el papel de la semejanza intergrupal y la identidad en las relaciones intergrupales.

La relación de la percepción de semejanza endoexogrupal y la identidad endogrupal con las actitudes hacia varios grupos de inmigrantes se analiza con una muestra integrada por 400 estudiantes canarios de Secundaria. Se les aplicó un cuestionario (autocumplimentado en el aula) que incluía tres escalas: identificación con el endogrupo, similitud intergrupal y una escala de *distancia social* (basada en la *escala de Bogardus*). Los participantes tenían que declarar el nivel de agrado hacia cuatro situaciones que implican distinto grado de proximidad con respecto a los miembros de los seis exogrupos sobre los que se preguntaba: inmigrantes de mayor presencia en las Islas Canarias (alemanes, ingleses, marroquíes, negros africanos, peninsulares y venezolanos). Los supuestos concretos sobre los que se pregunta son:

- Estudiar en el mismo colegio/instituto donde hay estudiantes marroquíes (o alemanes, venezolanos, etc.)
- Tener como profesor a...
- Tener como vecinos de la puerta de al lado a...
- Enterarnos de que una hermana está embarazada de...

Cada situación estaba seguida de una escala de 7 puntos, donde el 1 significa “muy poco agradable” y el 7 “muy agradable”. Del estudio se concluye (tras aplicar un análisis multidimensional de similitud intergrupal) que las personas hacia las que hay una mayor distancia social son los marroquíes, seguidos de los negros africanos, alemanes e ingleses.

2.4.2. Escala diferencial de Thurstone

En 1929, Louis Leon Thurstone propone, junto con E. J. Chave (en *The Measurement of Attitudes*. University of Chicago Press), un nuevo procedimiento para medir actitudes, que años después perfilaría en una publicación en solitario (1931: *The Measurement of Social Attitudes*. University of Chicago Press). La escala incluye una serie de proposiciones (favorables o contrarias) relativas a una determinada actitud, expresadas de manera categórica (como aseveraciones). De cada una se pide que se indique si se está de acuerdo o en desacuerdo. El promedio de las respuestas resume la actitud hacia el objeto de estudio. Para su elaboración se precisa:

- a) *Determinar la actitud a medir* (en consonancia con los objetivos y el marco teórico de la investigación).
- b) *Generar un número elevado de ítems* (de 100 a 150), que han de cubrir el continuo de la actitud que se mide: desde el extremo más positivo hasta el más negativo. No han de ser frases largas, ni contener dos o más ideas, ni dobles negativas, ni incluir referencias ambiguas o adverbios de cantidad (recomendaciones comunes en la elaboración de preguntas de encuesta como muestra el capítulo 6). Asimismo conviene que la ordenación de los ítems en la escala mantenga una estructura similar a la del cuestionario: que los primeros ítems sean los más suaves o menos definidores de la actitud; los que expresan una mayor controversia, en el segundo tercio de la escala; y los intermedios, al final.
- c) *Un grupo de expertos*, llamados “jueces” (preferiblemente 30 o más) *evalúan la serie de ítems*, indicando en qué punto del continuo de la actitud se sitúan. Para ello, cada juez o experto clasifica cada *ítem* en uno de los 11 montones que comprende el continuo de la actitud que gradúan: desde “bastante desfavorable” (1) hasta “bastante favorable” (11), siendo el “6” el valor neutro. Los ítems ubicados en montones diferentes quedan excluidos, por considerarse “ambiguos” para medir la actitud. Para ello puede elaborarse una *matriz* cuyas columnas incluyan las valoraciones que el conjunto de expertos da a cada ítem. Se calcula la *mediana* (también puede ser la *media*, aunque Thurstone prefería la *mediana*) de cada columna, que representa a cada ítem, y el *rango interpercentílico* ($P_{75} - P_{25}$). Éste es elegido por Thurstone para medir la dispersión discriminante de cada ítem. Se retendrá todo ítem cuya dispersión no supere el valor de “2” (que suscite consonancia en su valoración).
- d) *Se listan*, de forma aleatoria, los 20 o 30 ítems finalmente seleccionados para aplicarse a la *muestra* de estudio. El cuestionario no ha de incluir los valores escalares promedios dados a cada ítem en la *prueba de jueces*. Para cada ítem se ha de indicar si se está “de acuerdo” o “en desacuerdo”. La puntuación de cada sujeto (en la escala) se obtiene de la *media* (o *mediana*) de los valores escalares de los ítems en que ha dicho estar “de acuerdo”.

La laboriosidad que exige la elaboración de la *escala Thurstone* explica su uso limitado en la investigación social. A ello se suma que únicamente mide el acuerdo o el desacuerdo con aseveraciones, no la intensidad del acuerdo y el desacuerdo. Además, puede alcanzarse la misma puntuación conjunta con diferentes combinaciones de ítems. No obstante, la *escala Thurstone* puede combinarse con otras escalas (como la de *Likert*) y crear mediciones de tipo ordinal.

EJEMPLO DE ESCALA THURSTONE

Los siguientes ítems con sus *valores escalares* están tomados de una *escala de Thurstone* incluida en su publicación de 1931 (*The Measurement of Social Attitudes*), y reseñada por Lamberth (1980: 230). Es una escala de actitudes hacia los negros. Las frases o ítems elegidos concuerdan con la medición tradicional de racismo presente en aquellas fechas: el llamado *racismo tradicional o biológico*.

Valor escalar	Ítem		
10,3	– Creo que el negro tiene los mismos derechos sociales que el blanco	De acuerdo	_____
		En desacuerdo	_____
10,3	– El negro debería ser considerado como igual al hombre blanco y disfrutar de sus mismas ventajas	De acuerdo	_____
		En desacuerdo	_____
7,7	– El negro es perfectamente capaz de cuidar de sí mismo, si el hombre blanco le deja tranquilo	De acuerdo	_____
		En desacuerdo	_____
5,4	– No me interesa en absoluto la situación social del negro	De acuerdo	_____
		En desacuerdo	_____
2,7	– En ningún caso los niños negros deberían asistir a la misma escuela que los niños blancos	De acuerdo	_____
		En desacuerdo	_____
0,9	– El negro está siempre un poco por encima de los animales	De acuerdo	_____
		En desacuerdo	_____
0,9	– El negro debería ocupar el lugar más bajo entre los seres humanos	De acuerdo	_____
		En desacuerdo	_____

Los ítems figuran en el cuestionario de forma aleatoria, no siguiendo ningún orden (ascendente o descendente): se alternan ítems que expresan actitudes positivas con negativas y neutras. No figuran los *valores escalares* dados por los expertos (o *jueces*). Sólo la frase junto con las opciones de respuesta posibles: de acuerdo y en desacuerdo, que es marcada (con una cruz) por el encuestado. La *media* o *mediana* de los valores escalares atribuidos a los ítems representará la actitud que se tiene hacia los negros.

2.4.3. Escalograma o escala acumulativa de Guttman

Louis Guttman propuso en 1944 (“A basis for scaling qualitative data”. *American Sociological Review*, 9: 139-150) un procedimiento escalar acumulativo para medir actitudes. En él los ítems se ordenan de tal manera que la respuesta positiva a uno de ellos suponga también acuerdo con los ítems de orden de rango inferior (que se sitúan debajo de él). Ello exige una clasificación jerárquica y acumulativa de los ítems.

A diferencia de la *escala de Thurstone*, el universo de ítems se reduce a 30 e inclusive menos, si aparecen acompañados de otras preguntas en el cuestionario. Además, se elimina la *prueba de jueces*. Basta su previa comprobación en una muestra de la población: la prueba piloto o *pretest* del cuestionario.

Las categorías de respuesta pueden ser *dicotómicas* (“de acuerdo” 1, “en desacuerdo” 0; “sí” 1, “no” 0) o incluir más de dos opciones de respuesta como en la *escala Likert*. En este último caso, la puntuación más elevada se asigna al valor de la respuesta que sea más favorable a la actitud. *Por ejemplo*, “completamente de acuerdo” (4), “de acuerdo” (3), “indiferente” (2), “en desacuerdo” (1), “completamente en desacuerdo” (0). En ambos casos se sigue el mismo procedimiento de escalamiento, aplicándose la técnica de *escalograma* llamada *Cornell* (el nombre de la universidad donde Guttman trabajaba por aquella fecha), pero con cinco categorías de respuesta.

EJEMPLOS DE ESCALOGRAMA GUTTMAN

- A) De Vellis (1991) ilustra el *escalograma* de Guttman en una *escala de aspiraciones de los padres* hacia el logro educativo de sus hijos:

Lograr el éxito escolar es la única forma de que mis hijos compensen mis esfuerzos como padre	De acuerdo	1
	En desacuerdo	0
Ir a un buen colegio y obtener un buen trabajo son importantes para la felicidad de mis hijos	De acuerdo	1
	En desacuerdo	0
La felicidad es más probable, si una persona ha logrado sus metas educativas y materiales	De acuerdo	1
	En desacuerdo	0

En una *escala Thurstone* los ítems quedarían enunciados de esta forma:

Lograr éxito es sólo una forma de que mis hijos compensen mis esfuerzos como padre	De acuerdo	_____
	En desacuerdo	_____
Ir a un buen colegio y obtener un buen trabajo son importantes, pero no esenciales en la felicidad de mis hijos	De acuerdo	_____
	En desacuerdo	_____
La felicidad nada tiene que ver con lograr metas educativas o materiales	De acuerdo	_____
	En desacuerdo	_____

- B) Propuesta de *escalograma de Guttman* de Rafael García Pérez (2004: “Diagnóstico en la teleformación: construcción y validación de un escalograma Guttman”. *Revista de Investigación Educativa*, vol. 22, n.º 1: 277-302) para evaluar el nivel de dominio de las técnicas e instrumentos de diagnóstico en educación de estudiantes universitarios. Se diferencia entre niveles de “principiante”, “iniciado” y “experto” para la autoselección de actividades de aprendizaje:

	Sí	No
– Usted conoce este instrumento o estrategia de diagnóstico y evaluación educativa	1	0
– Usted sabe aplicar esta técnica de diagnóstico en un caso o situación concreta	1	0
– Usted puede identificar y explicar los elementos/pasos fundamentales que la componen	1	0
– Usted puede elaborar un material educativo sobre su naturaleza y uso en educación ...	1	0
– Usted conoce las teorías que fundamentan su aplicación en el campo educativo	1	0
– Usted podría formar al profesorado o a técnicos sobre el uso de esa técnica en educación	1	0
– Usted puede plantear argumentaciones técnicas y científicas sobre su uso en educación	1	0
– Usted conoce las líneas de investigación y los últimos avances sobre los enfoques y teorías científicas que fundamentan ésta técnica	1	0

A la escala acompaña la siguiente nota para su uso: “Responda positiva o negativamente a las preguntas aplicándolas a la técnica e instrumento de diagnóstico cuyo conocimiento desee autoevaluar. El primer escalón respondido negativamente indica el nivel del usuario; los dos primeros escalones corresponden al nivel de “principiante”, los escalones 3, 4 y 5 son del nivel de “iniciado”, y los ítems 6, 7 y 8 corresponden a una problemática educativa vinculada al nivel de “experto”.

Las respuestas obtenidas en la escala se colocan en una *matriz*. En las columnas se sitúan las preguntas, especificándose la respuesta: “sí” (1), “no” (0). En las filas, los sujetos que contestan. Para cada sujeto, se suma el número de respuestas afirmativas. Se cambia el orden de las filas de la matriz original para que los sumatorios más elevados ocupen las primeras posiciones y los más bajos las últimas. Los ítems en las columnas también se reordenan para que en las primeras columnas aparezcan los ítems cuyo sumatorio sea más bajo y en la última el más alto, como ilustra el cuadro 2.9.

Después se calcula el *índice de reproductividad* para validar la escala:

$$R=1-\frac{E}{Q \times S}$$

Donde: E = número total de errores

Q = número de ítems

S = número de sujetos que responden

La *escala* se considera “válida,” cuando el *índice de reproductividad* es mayor o igual a “0,90”. En caso de invalidez (un índice inferior a 0,90), se procede a la eliminación gradual de los ítems que presentan un mayor número de errores, hasta que el *índice* supere el valor de referencia que concede *validez* a la escala: “0,90”. Por lo que se eliminan aquellos ítems que han obtenido respuestas de mayor dispersión.

CUADRO 2.9. Escalograma de seis ítems con respuesta favorable (1) y desfavorable (0)

Sujetos que contestan	Ítems						Puntuación de los sujetos
	3	6	1	2	5	4	
1	1	1	1	1	1	1	6
2	1	1	1	1	1	1	6
3	1	1	1	1	1	1	6
4	1	1	1	0*	1	1	5
5	0	1	1	1	1	1	5
6	0	1	1	1	1	1	5
7	0	0	1	1	1	1	4
8	0	0	1	1	1	1	4
9	0	0	1	1	0*	1	3
10	0	1*	0	1	1	1	3
11	1*	0	0	0	1	1	2
12	0	0	0	0	0	1	1
13	0	0	0	1*	0	0	1
Puntuación categorías	5	7	9	10	10	12	53
Errores*	1	1	0	2	1	0	5

* Errores: Valores “1” ubicados en la zona de “0” y a la inversa, “0” en la zona de “1”.

EJEMPLO DE APLICACIÓN DEL ÍNDICE DE REPRODUCTIVIDAD EN LA ELABORACIÓN DE UNA ESCALA DE GUTTMAN

González Blasco (1989/2000) ilustra la confección de una *escala de Guttman* con cinco proposiciones (o ítems) extraídas de una escala utilizada por C. Srole en la medición del *nivel de anomia* (en García Ferrando, M. (1971): “Estructura comunitaria y desarrollo organizado”. *Revista de Estudios Sociales*, 3: 3-18).

Las respuestas se rotaron, de forma que su ordenación minimizara los *errores*, según los criterios de Guttman. Con los cinco ítems se obtuvo un *escalograma*. Mediante un programa informático se calcularon los *errores* de cada ítem, sus porcentajes de respuesta y el *coeficiente de reproductividad*. Este último resultó ser 0,8767. Un valor algo inferior a 0,90. Por lo que, se opta por eliminar el ítem de mayor error (54,00), el número 4. Tras

su eliminación, se repite el proceso con los cuatro ítems restantes. Como el nuevo *índice de reproductividad* es 0,9034, la *escala* integrada por las cuatro proposiciones restantes queda finalmente aceptada.

CUADRO 2.10. Escala para la medición del nivel de anomia

Ítems	% respuesta afirmativa	Error	% Error
– No es útil escribir a los funcionarios públicos porque a menudo no están interesados en los problemas del hombre de la calle.	67,3	30,50	7,90
– Actualmente hay que vivir al día y dejar que el mañana se cuide a sí mismo.	54,4	50,00	12,95
– A pesar de lo que se dice, las cosas están cada vez peor para el hombre medio y no mejor.	41,7	51,00	13,21
– El traer niños a este mundo es un problema, dado cómo se presenta el futuro.	40,7	54,00	13,99
– En estos días uno no sabe en quién confiar.	35,2	52,50	13,60
Total casos observados: 443 residentes del Estado de Mississippi			

2.4.4. Escala aditiva de Likert

En 1932 Rensis Likert propone (en “A technique for the measurement of attitudes”. *Archives of Psychology*, 140: 44-53) uno de los procedimientos escalares más aplicados en la investigación social. Presenta las siguientes peculiaridades, que le distinguen de otras escalas:

- No precisa un número de ítems iniciales tan elevado como la escala Thurstone.* Basta con reunir el doble o el triple de los que se desea incluir en la escala final. Es habitual proponer circa 60 ítems para, después, quedarse con 20-25.
- La escala ha de ser equilibrada:* ha de contener un número similar de ítems claramente favorables como desfavorables hacia el objeto que se mide.
- No deberían incluirse aseveraciones “muy suaves”,* porque suscitan demasiado acuerdo, haciéndolas poco atractivas o interesantes a efectos analíticos. Por la misma razón, tampoco deberían incluirse ítems que despierten acuerdo unánime, por ser poco definidores de la actitud que se mide.
- Es una escala politómica,* generalmente con cinco categorías de respuesta para cada ítem. *Por ejemplo*, “muy de acuerdo”, “de acuerdo”, “ni de acuerdo ni en desacuerdo”, “en desacuerdo”, “muy en desacuerdo”. Las puntuaciones (1, 2, 3, 4, 5 o, a la inversa, 5, 4, 3, 2, 1) se asignan en consonancia con el significado de la respuesta para la actitud que miden. Asimismo pueden apli-

carce los códigos numéricos +2, +1, 0, -1, -2. Esta opción tiene la ventaja de que el “0” implica neutralidad o ambigüedad completa, mientras que un número negativo alto significa una actitud que se opone a la opinión representada por un número positivo alto. La puntuación global de la escala suele obtenerse “sumando” (al ser una escala *sumativa* o aditiva) todas las puntuaciones registradas (incluidas las inversas).

EJEMPLO DE PUNTUACIÓN DE CATEGORÍAS EN UNA ESCALA LIKERT

- *Los inmigrantes legalmente establecidos en España deberían tener los mismos derechos sociales que los españoles*

5	4	3	2	1
Muy de acuerdo	De acuerdo	Indiferente	En desacuerdo	Muy en desacuerdo

- *Todos los inmigrantes, legales o ilegales, deberían ser repatriados a sus países de origen*

1	2	3	4	5
Muy de acuerdo	De acuerdo	Indiferente	En desacuerdo	Muy en desacuerdo

Las respuestas se combinan de manera que aquellas que supongan una actitud más favorable obtendrán una puntuación más elevada, mientras que a la actitud más desfavorable le corresponderá la más baja. En este ejemplo podría haberse aplicado, igualmente, las puntuaciones de +2, +1, 0, -1, -2, al primer ítem y, a la inversa, en el segundo. El primer ítem connota una actitud positiva hacia la inmigración, mientras que el segundo, negativa. Por esta razón se les adscribe valores numéricos opuestos, aunque en la fase de análisis. En el cuestionario las alternativas de respuesta iguales han de figurar con el mismo código numérico para evitar influir en la respuesta. Tras la recogida de la información, es cuando los códigos numéricos se modifican en la dirección de la actitud que midan.

Respecto a las *categorías de respuestas* no está claro el número ideal. Sheatsley (1983) desaconseja más de cinco porque algunas personas pueden tener dificultades para encontrar las diferencias de matiz entre unas respuestas y otras. Para Krosnick y Fabrigar (1997) la longitud de la escala óptima se sitúa entre cuatro y siete puntos. Además constatan que la *fiabilidad* es mayor en escalas de siete a nueve puntos en investigaciones longitudinales y de cinco a siete puntos en investigaciones transversales (o seccionales). Por otra parte, está el debate sobre la inclusión de categorías intermedias (“ni de acuerdo ni en desacuerdo”, “indiferente”) porque favorece su elección en

personas que no quieren definirse o pensar en una respuesta (como asimismo se verá en el capítulo 6). También la cuestión de si etiquetar todos los puntos de la escala con palabras o sólo los extremos (y los demás con números). Incluir etiquetas en todos los puntos de la escala puede ayudar a clarificar su significado y a la introducción de matices. Principalmente, en poblaciones de bajo nivel cultural.

- e) *No precisa de la prueba de jueces* (al igual que el *escalograma de Guttman*). Puede administrarse a una pequeña muestra representativa (de la población de estudio): entre 50-100 personas que han de cubrir el rango de actitudes que se desea detectar con la escala.
- f) La información que aporten se utilizará para *seleccionar los ítems* finalmente a incluir en la escala definitiva: los menos ambiguos y más discriminantes para medir la actitud. Se calcula la puntuación total de cada individuo en la escala, a partir de sus puntuaciones en cada ítem. Aquellos que presenten una actitud más favorable a la inmigración deberán tener una puntuación más alta; los más contrarios o reacios a la inmigración, la más baja.

Los sujetos se ordenan en orden decreciente, de acuerdo con su puntuación total. La primera posición la ocupa el de puntuación más elevada, mientras que la última, el de menor puntuación. Con el 25% de los sujetos con puntuaciones más elevadas (grupo alto) y el 25% con las más bajas (grupo bajo), se analiza la adecuación de cada ítem para medir la actitud. Puede seguirse alguno de los siguientes procedimientos:

- *Calcular la media de cada ítem en ambos grupos* (alto y bajo) y sus *desviaciones típicas* correspondientes. Su significatividad se comprueba mediante la prueba estadística *t* de *Student*. Los ítems a los que corresponda un valor *t* empírico inferior al teórico, al nivel de probabilidad elegido (normalmente un nivel de significación de 0,05, que equivale a una probabilidad de acierto del 95%), serán descartados por no ser discriminantes: no presentan diferencias estadísticamente significativas entre ambos grupos.
- *Calcular el coeficiente de correlación de cada ítem con el total de la escala*. Los ítems elegidos serán aquellos que presenten una correlación estadísticamente significativa (normalmente, una correlación superior a 0,35).
- *Calcular la mediana de cada ítem en ambos grupos y elaborar una tabla de contingencia*, con dos filas (una de puntuaciones por debajo de la mediana; la segunda de puntuaciones por encima de la mediana) y dos columnas (grupo alto, grupo bajo). El poder discriminante de cada ítem se comprueba mediante el *test de chi-cuadrado* o el *test de Fischer*. Como en toda prueba de significatividad, cuando los valores empíricos superen los correspondientes valores teóricos o críticos, significa que el ítem es válido para diferenciar los grupos (tiene poder discriminante).

- *El poder discriminatorio de los ítems puede igualmente comprobarse mediante el análisis discriminante.* Éste ayuda a la obtención de los ítems que más definen a los integrantes de cada grupo y los que más diferencian a unos grupos de los otros (los más discriminantes, como se verá en el capítulo 8).

Otra técnica multivariable de gran utilidad es el *análisis factorial*, tanto en la fase exploratoria como en la confirmatoria. Permite comprobar si la escala está compuesta por ítems que miden dos o más dimensiones de la actitud y cómo deberían ponderarse para componer las escalas.

Una vez comprobado el poder discriminatorio de cada ítem, se retienen aquellos que muestren ser estadísticamente significativos en la diferenciación de ambos grupos (alto y bajo). Se añade el requisito de que la mitad de los ítems midan una actitud favorable y la otra mitad (aproximadamente) una desfavorable. El diseño de la escala concluye listando los ítems en orden aleatorio, intentando que los situados en el primer tercio sean los menos definidores de la actitud o más suaves, los ubicados en el segundo tercio los más controvertidos, y en el tercio final los intermedios. Los ítems que representen actitudes diferentes se alternan para evitar el *sesgo de aquiescencia* (o tendencia al “acuerdo” indistintamente del contenido del ítem). Este es un *sesgo* muy frecuente cuando los primeros ítems indican una misma actitud o consonancia de significado.

Es habitual retener en torno a veinte ítems. Si la escala se anexiona a otras preguntas en un cuestionario, se aconseja reducir el número de ítems a los que más definan la actitud. Como advierte Sheatsley (1983: 210), “cuando el número de ítems excede de 8 o 10, muchos respondientes se aburrirán y responderán sin propósito, sin pensamiento real. Simplemente para pasar la lista tan rápido como sea posible”.

EJEMPLO DE ESCALA LIKERT

- A) *Escala Likert* aplicada por Reynaldo Alarcón en 2006 (“Desarrollo de una escala factorial para medir la felicidad”, *Revista Interamericana de Psicología/Interamerican Journal of Psychology*, vol. 40, n.º 1: 99-106) para medir la felicidad.

La “*felicidad*” se define como un estado afectivo de satisfacción plena que experimenta subjetivamente el individuo en posesión de un bien anhelado. Para su medición se desarrolla una escala cuya *validez* y *fiabilidad* se comprueba previo a su administración a una muestra de 709 estudiantes de universidades públicas y privadas de Lima (Perú).

Tras la definición operacional del constructo *felicidad* y sus indicadores, se elaboraron 52 ítems, unos propios y otros procedentes de escalas aplicadas en investigaciones previas (E. Diener, R. E. Emmons, R. J. Larsen y S. Griffin, 1985: “The satisfaction with life scale”, *Journal of Personality Assessment*, 49, pp. 71-75; P. Hills y M. Argyle, 2002: “The Oxford Happiness Questionnaire: A compact scale for the measurement of psychological well-being”, *Personality and Individual Differences*, 33, pp. 1073-1082). Los ítems fueron sometidos a evaluación de cinco jueces, que dieron sugerencias para eliminar o reformular varios de ellos.

Con los ítems “aprobados” se organizó un inventario experimental de 42 reactivos, el mismo que se administró a una *muestra* de 80 estudiantes universitarios. El análisis de ítems del estudio piloto eliminó 12 reactivos. Quedaron 30 ítems, que pasaron a componer la *escala de felicidad*. Una escala tipo *Likert*, de cinco alternativas de respuesta: “totalmente de acuerdo”, “acuerdo”, “ni acuerdo ni desacuerdo”, “desacuerdo” y “totalmente en desacuerdo”. Al extremo positivo correspondía 5 puntos; al negativo, 1 punto. Los ítems fueron redactados para que denotaran un posicionamiento favorable o desfavorable a la *felicidad*, y se colocaron de forma aleatoria en el cuestionario. El análisis factorial de *componentes principales* (y rotación *varimax*) mostró que la *felicidad* es un comportamiento multidimensional, integrado por cuatro factores (que se detallan a continuación):

- Factor 1. *Sentido positivo de la vida*: sus ítems indican ausencia de estados depresivos, pesimismo y de vacío existencial.
- Factor 2. *Satisfacción con la vida*: los reactivos de este factor trasuntan estados subjetivos de satisfacción por la posesión de los bienes deseados.
- Factor 3. *Realización personal*: expresa felicidad plena y no estados temporales de “estar feliz”; también, autosuficiencia.
- Factor 4. *Alegría de vivir*: refiere lo maravilloso que es la vida, las experiencias positivas de la vida y sentirse generalmente bien.

CUADRO 2.11. Matriz de factores extraídos por rotación *varimax* y cargas factoriales de los ítems

ÍTEMS	F1	F2	F3	F4
26. Todavía no he encontrado sentido a mi existencia	.68			
20. La felicidad es para algunas personas, no para mí	.68			
18. Tengo problemas tan hondos que me quitan la tranquilidad	.67			
23. Para mí, la vida es una cadena de sufrimientos	.67			
22. Me siento triste por lo que soy	.66			
02. Siento que mi vida está vacía	.66			
19. Me siento un fracasado	.64			
17. La vida ha sido injusta conmigo	.62			
11. La mayoría del tiempo no me siento feliz	.62			
07. Pienso que nunca seré feliz	.59			
14. Me siento inútil	.53			
04. Estoy satisfecho con mi vida		.68		
03. Las condiciones de mi vida son excelentes		.67		
05. La vida ha sido buena conmigo		.66		

[.../...]

CUADRO 2.11. (Continuación)

ÍTEMS	F1	F2	F3	F4
01. En la mayoría de las cosas, mi vida está cerca de mi ideal		.57		
06. Me siento satisfecho con lo que soy		.55		
10. Me siento satisfecho porque estoy donde tengo que estar		.44		
24. Me considero una persona realizada			.67	
21. Estoy satisfecho con lo que hasta ahora he alcanzado			.64	
27. Creo que no me falta nada			.61	
25. Mi vida transcurre plácidamente			.58	
09. Si volviera a nacer no cambiaría casi nada de mi vida			.46	
08. Hasta ahora he conseguido las cosas que para mí son importantes			.41	
12. Es maravilloso vivir				.66
15. Soy una persona optimista				.61
16. He experimentado la alegría de vivir				.58
13. Por lo general, me siento bien				.49
Coeficientes Alfa	.88	.79	.76	.72

B) *Escala Likert* aplicada en el estudio de Marta Torres Rius, Francisco Alvira, Francisca Blanco y M. Sandi Pérez (1994: *Relaciones padres/hijos*. Ministerio de Asuntos Sociales. Colección *Estudios*), y que adopta el siguiente formato en el cuestionario del estudio:

P. 34. A continuación voy a leerle una serie de frases. Me gustaría que me dijera, para cada una de ellas, si está muy de acuerdo, de acuerdo, indiferente, en desacuerdo o muy en desacuerdo.

Frases	Muy de acuerdo	De acuerdo	Indiferente	En desacuerdo	Muy en desacuerdo	NS/NC	
Preferiría no haber tenido hijos	1	2	3	4	5	9	(22)
Los hijos han sido una gran satisfacción en mi vida	1	2	3	4	5	9	(23)
Hoy en día los hijos no respetan a los padres	1	2	3	4	5	9	(24)
Si se explican las cosas cualquier niño entiende las razones de su padre	1	2	3	4	5	9	(25)
Los hijos vienen con un pan debajo del brazo	1	2	3	4	5	9	(26)
No hay cosa que más satisfaga a un padre que la sonrisa de su hijo	1	2	3	4	5	9	(27)

[.../...]

<i>Frases</i>	<i>Muy de acuerdo</i>	<i>De acuerdo</i>	<i>Indiferente</i>	<i>En desacuerdo</i>	<i>Muy en desacuerdo</i>	<i>NS/ NC</i>	
Los padres son los que han de tomar exclusivamente las decisiones y los hijos obedecerlas	1	2	3	4	5	9	(28)
No creo que pegar sea el sistema de enseñar nada	1	2	3	4	5	9	(29)
Cualquier decisión que se tome, si ésta es importante, debe discutirse con todos incluidos los hijos	1	2	3	4	5	9	(30)
La disciplina en educación es la clave del éxito	1	2	3	4	5	9	(31)
La letra con sangre entra	1	2	3	4	5	9	(32)
A los niños desde muy pequeños hay que acostumbrarlos a razonar	1	2	3	4	5	9	(33)
Me basta con que mis hijos se críen con salud	1	2	3	4	5	9	(34)
Los hijos son un problema que nunca acabas de llevar	1	2	3	4	5	9	(35)
Por más que uno se esfuerce, al final los hijos salen como quieren	1	2	3	4	5	9	(36)

Con estos 15 ítems se mide la actitud de los padres hacia los hijos. La escala incluye ítems que denotan una actitud tanto positiva (*por ejemplo*, “Los hijos han sido una gran satisfacción en mi vida”), como negativa (“Preferiría no haber tenido hijos”) o indiferente (“Me basta con que mis hijos se críen con salud”). Las opciones de respuesta mantienen en el cuestionario el mismo código numérico en todos los ítems. En la fase de análisis los códigos se recodifican para que estén en consonancia con la actitud que miden y se aplica un *análisis factorial de componentes principales*. De él se extraen tres modelos de relación padres-hijos: modelo autoritario (frases 1, 7 y 10), de *laissez-faire* (frases 13, 14 y 15) e inductivo-de apoyo (frases 4, 6, 9 y 12). Los cinco ítems restantes no mostraron utilidad en la clasificación de las relaciones padres-hijos.

Un *problema* importante de la *escala Likert* es que la misma puntuación global puede obtenerse de diferentes combinaciones de respuesta (Procter, 1993). A ello se suma otra crítica extensible a otras escalas que miden las actitudes a través de declaraciones verbales: la persona puede descubrir qué actitud se mide y modificar su respuesta (si quiere ajustarse a las demandas del investigador o a aquellas que

perciba como socialmente deseables). Por ello son vulnerables al *sesgo de deseabilidad social* (del que se hablará en los capítulos 5 y 6).

2.4.5. Diferencial semántico de Osgood

Propuesto por Charles E. Osgood, George J. Suci y Percy H. Tannenbaum en 1957 (*The measurement of meaning*. University of Illinois Press), como un procedimiento para asegurar el significado connotativo de los conceptos. Muestra gran utilidad para la medición de *estereotipos* y, en general, de actitudes hacia diferentes grupos sociales. De manera especial, en personas con fuertes reacciones emocionales, como señalan Henerson, Morris y Fitz-Gibbon (1987). Para Lamberth (1980: 231) constituye “un instrumento útil y flexible de medición”. No existe un modelo único de *diferencial semántico*. El tema y los adjetivos antónimos o polares elegidos hacen que las escalas sean muy diversas.

El *diferencial semántico* parte del supuesto de que se puede determinar los sentimientos (positivos o negativos) de las personas, hacia un objeto concreto de actitud, mediante su posición ante adjetivos polares o antónimos. El objeto de actitud a medir se sitúa en el cabecero de la pregunta. A la persona se le pide que marque una de las posiciones (generalmente siete) que median entre dos adjetivos antónimos. Para su elaboración se recomienda que los pares de adjetivos (que expresan dimensiones del concepto a medir) se alternen, de forma aleatoria, para que las respuestas positivas (o negativas) no caigan en el mismo extremo. *Por ejemplo*:

En su opinión, ¿cómo son los marroquíes? Coloque una X en una de las siete posiciones comprendidas entre cada par de adjetivos polares.

MARROQUÍES

Sinceros	—	—	—	—	—	—	Falsos
Holgazanes	—	—	—	—	—	—	Trabajadores
Honestos	—	—	—	—	—	—	Deshonestos
Sucios	—	—	—	—	—	—	Limpios
Activos	—	—	—	—	—	—	Pasivos
Hostiles	—	—	—	—	—	—	Amigables
Tolerantes	—	—	—	—	—	—	Intolerantes
	1	2	3	4	5	6	7

Los adjetivos pueden ser *bipolares* (por ejemplo, amigable-hostil) o *unipolares*. Estos últimos indican la existencia o ausencia de un único atributo. *Por ejemplo*, amigable-no amigable. La puntuación de una persona se obtiene asignando “1” a la posición que expresa la respuesta más negativa; el “7”, a la más positiva; y las intermedias, desde 2 a 6, consecutivamente. La puntuación global será el promedio de todas las respuestas para la totalidad de los enunciados propuestos. Si en el *pretest*

se obtuviera que un porcentaje elevado de respuestas se sitúe en la zona neutra o intermedia (próxima al 4), debería procederse a la sustitución de los pares de adjetivos antónimos por otros que muestren mayor utilidad para medir el objeto de interés.

EJEMPLO DE DIFERENCIAL SEMÁNTICO DE OSGOOD

Diferencial semántico del envejecimiento desarrollado por Feliciano Villar Posada en 1997 [“Construcción y evaluación en diferentes cohortes del DSE (Diferencial semántico del envejecimiento)”, *Anales de Psicología*, vol. 13, n.º 1: 31-37].

El objetivo del estudio era construir una escala que midiera las actitudes hacia el *envejecimiento*. Contó con la colaboración voluntaria de 180 personas que en ese momento vivían en Barcelona o su área metropolitana. El muestreo fue de tipo intencional, teniendo en cuenta dos criterios: sexo y edad. La *escala* debía ser lo suficientemente breve y sencilla como para que pudieran completarla de forma individual personas con un mínimo nivel educativo. En este sentido, la técnica del *diferencial semántico* parecía especialmente adecuada; además de ser frecuentemente utilizada para medir actitudes hacia la vejez o hacia las personas mayores. Para la selección de ítems se realizó una prueba piloto. Se reunieron 31 pares de adjetivos bipolares (a partir de los utilizados en otras escalas para evaluar actitudes hacia las personas mayores).

El continuo que se definió entre los adjetivos bipolares fue de siete intervalos. *Por ejemplo*, en el par *fuerte-débil*, los sujetos podían responder si el envejecimiento era muy fuerte, bastante fuerte, algo fuerte, ni fuerte ni débil, bastante débil, algo débil o muy débil. Los diferentes intervalos se puntuaron desde “1” (actitud más negativa) hasta “7” (actitud más positiva) en cada par de adjetivos, estableciéndose el punto neutro de actitud en la puntuación “4”. Los pares de adjetivos se presentaron aleatorizando la dirección del continuo; es decir, a veces éste iba del polo positivo al negativo (por ejemplo, *bonito-feo*), mientras otras veces del negativo al positivo (por ejemplo, *improductivo-productivo*). La *muestra* del estudio piloto estuvo compuesta por 82 personas, de las que 44 eran mujeres y 38 varones, con edades comprendidas entre 20 y 56 años. La edad media era 31,1 años. La extracción social de la *muestra* era prioritariamente de nivel medio o medio bajo. Para su realización, se dieron las siguientes indicaciones:

En la siguiente página encontrará una serie de pares de adjetivos de significado opuesto que pueden ser utilizados para describir el proceso de envejecimiento. Su tarea consistirá en valorar en cada par **cuál de los dos adjetivos opuestos describe mejor el proceso de envejecimiento**. Para ello ha de utilizar las 7 rayas que hay entre un adjetivo y otro, y colocar un aspa (x) en la que representa mejor su opinión. Recuerde que no hay respuestas correctas o incorrectas y ponga una única cruz expresando su opinión en TODAS las parejas de adjetivos.

En la parte superior de la hoja en la que figuraban los pares de adjetivos se colocó, con un tipo de letra resaltada, la palabra ‘envejecimiento’ con el fin de centrar la atención de la persona en el concepto que debía evaluar. De los 31 ítems de los que constaba la

escala, 18 superaron los criterios de aceptación: 1) tener una variabilidad suficiente y en ningún caso inferior a un tercio de la media (para discriminar entre personas con actitudes positivas y negativas); 2) presentar una correlación entre el ítem y la puntuación total obtenida en toda la escala no inferior a 0,4.

Como paso previo a los análisis, las puntuaciones de los pares de adjetivos se transformaron de manera que en todos ellos una puntuación “1” correspondiese a una respuesta extrema en el polo negativo o indeseable del par de adjetivos, y una puntuación “7” correspondía a una puntuación extrema en el polo positivo o deseable. Las respuestas se sometieron a un análisis de *componentes principales* (rotación varimax). El cuadro 2.12 especifica el peso que presentaban los ítems para cada uno de los componentes.

CUADRO 2.12. Matriz de componentes extraídos por rotación varimax y cargas factoriales de los ítems

ÍTEMS	C1	C2	C3
Malo - Bueno	.81		
Penoso - Grato	.79		
Feo - Bonito	.78		
Negativo - Positivo	.71		
Inoportuno - Oportuno	.62		
Desagradable - Agradable	.61		
Resignado - Esperanzado	.60		
Pesimista - Optimista	.55		
Enfermo - Sano		.73	
Débil - Fuerte		.72	
Pasivo - Activo		.64	
Inútil - Útil		.56	
Aburrido - Emocionante		.52	
Improductivo - Productivo		.51	
Hostil - Amistoso			.82
Intolerante - Tolerante			.70
Triste - Alegre			.51
Conservador - Progresista	—	—	—
<i>Porcentaje de varianza</i>	40,7	8,6	6,7

La actitud hacia el envejecimiento se sustenta básicamente en torno a una gran dimensión (que por sí sola da cuenta de algo más de un 40% de la varianza). Esta gran dimensión corresponde a un componente principal genérico de evaluación del envejecimiento, ya que los ítems que más cargan en ella son algunos como *malo-bueno*, *penoso-grato*, *feo-bonito* o *negativo-positivo*. De modo que las personas que puntúan bajo en estos pares tienden a contemplar el envejecimiento en términos negativos, como un proceso de declive y pérdida, mientras que aquellos que puntúan alto en estos pares contemplan el envejecimiento en términos positivos, como un proceso deseable.

A gran distancia de esta dimensión se encuentran las otras dos. El segundo componente principal comprende pares como *enfermo-sano*, *débil-fuerte* o *pasivo-activo*, especialmente relacionados con una visión del envejecimiento como un proceso bien caracterizado por el dinamismo y el vigor, bien como un proceso de estancamiento y pasividad. El tercero de los componentes principales, con cargas especialmente altas de los pares *hostil-amistoso* o *intolerante-tolerante*, es de difícil interpretación, aunque parece sugerir una visión del envejecimiento como un proceso amenazante y rígido o, por el contrario, susceptible de cambio.

Se comprueban las cualidades psicométricas de la escala. El coeficiente de *fiabilidad*, estimado mediante el estadístico *alfa de Cronbach*, alcanzó un valor de 0,91 (un nivel óptimo de *fiabilidad*). La correlación de los ítems con la puntuación total en la escala también es elevada, yendo desde un valor inferior de “0,43” (par *pesimista-optimista*) hasta un límite superior de “0,77” (par *intolerante-tolerante*), con una media de “0,62”. Lo que lleva al autor a concluir que, considerando los tres componentes principales, el *diferencial semántico de envejecimiento* se aproxima a las tres dimensiones de actitud que se supone mide todo *diferencial semántico*: *Evaluación* (componente principal I), *Actividad* (componente principal II) y *Potencia* (componente principal III).

2.5. La calidad de la medición: fiabilidad y validez

Una medición se considera de “calidad” cuando satisface criterios de *validez* y *fiabilidad*. En el ejemplo anterior, y en algunos precedentes, se han destacado criterios habituales para su comprobación, caso del coeficiente de *fiabilidad alfa de Cronbach*, aunque no es el único. El cuadro 2.13 sintetiza criterios estándares de *validez* y *fiabilidad*, que se desglosan en sus apartados correspondientes. La *validez* hace referencia a la correspondencia de los indicadores con los conceptos a medir, mientras que la *fiabilidad* se refiere a la consistencia de la medición.

CUADRO 2.13. Criterios de validez y de fiabilidad de la medición

VALIDEZ	FIABILIDAD
De constructo	Método <i>test-retest</i>
– Convergente	– Coeficiente de correlación
– Divergente	– Índice de inconsistencia
De contenido	Método <i>alternativo</i>
De criterio	Método de <i>las dos mitades</i>
– Concurrente	Método de la <i>consistencia interna alfa</i>
– Predictiva	de <i>Cronbach</i>

2.5.1. Validez de la medición

Toda evaluación del proceso de investigación parte del cumplimiento de criterios de *validez*. Se aplican al *diseño* de la investigación (como ya se viera en el capítulo 1), pero también al *instrumento* o procedimiento de medición aplicado. Respecto a la *medición*, ha de haber correspondencia entre el *concepto* teórico a medir y los *indicadores* (o variables empíricas) seleccionadas para su medición. Ha de comprobarse si se cubre toda su *dimensionalidad* y si los *indicadores* realmente “indican” lo que se pretende que indiquen; si proporcionan mediciones “válidas” del constructo o *concepto*. Es el primer criterio a valorar en la evaluación de la calidad de la medición. Una vez asegurada, se procede a evaluar su *fiabilidad*. Si bien, como señala Neuman (2000: 164), “la validez es más difícil de alcanzar que la fiabilidad”.

Carmines y Zeller (1979) diferencian tres modalidades básicas de *validez*:

- a) *De constructo*: si la medida utilizada coincide con aquella que teóricamente habría que esperar, a partir de las hipótesis que se derivan del marco teórico de la investigación.

Por ejemplo, en la medición de la *delincuencia* en encuestas de autodenuncia (a la población en general), la incidencia de la variable clase social en la explicación de la actividad delictiva se debilita. Al contrario de lo que sucede, cuando la encuesta se hace a la población recluida en centros penitenciarios. Por lo que habrá que precisar qué medida de “delincuencia” es más válida.

- b) *De contenido*: grado al que una medición empírica cubre la variedad de significados incluidos en el concepto; si consigue cubrir el “dominio” de significados del concepto.

Por ejemplo, en la medición de la *xenofobia* comprobar si los distintos indicadores presentes en una encuesta de actitudes hacia la inmigración abarcan toda la dimensionalidad del concepto que se mide. Ello exige la previa revisión bibliográfica de cómo se ha medido la *xenofobia* en diferentes estudios, las *dimensiones* que se estiman básicas para cubrir el dominio de su contenido.

- c) *De criterio*: la validez de la medición se comprueba comparándola con algún “criterio” (o medida generalmente aceptada) que se haya empleado con anterioridad para medir el mismo concepto. Dentro de ella se distingue entre:

- *Validez concurrente*: cuando la medición nueva se correlaciona con un criterio adoptado para el mismo momento. El *coeficiente de correlación* es el estadístico de uso más extendido para su cuantificación.

Por ejemplo, comparar los resultados de una encuesta de intención de voto (a pie de urna) con los “oficiales” al término de las elecciones.

- *Validez predictiva*: cuando la comparación se hace respecto a un “criterio” futuro que esté correlacionado con la medida.

Por ejemplo, comparar las respuestas recabadas en una encuesta de actitudes ante la inmigración a empresarios con la conducta que éstos posteriormente manifiestan cuando contratan a sus empleados.

Un instrumento de medida tiene *validez predictiva* cuando puede utilizarse para “predecir” una conducta futura. Al igual que la *validez concurrente*, la *predictiva* se cuantifica mediante el cálculo de coeficientes de *correlación* de las respuestas recabadas en la encuesta con medidas posteriores de su conducta que se consideran “válidas”.

Tanto la *validez concurrente* como la *predictiva* requiere la comparación de medidas alternativas de un mismo *concepto*. Cuanta más correlación haya entre ellas, mayor será su *validez*. Pero, como bien precisa De Vaus (1990), estos procedimientos de comprobación de la *validez* no están exentos de polémica. Presentan dos *problemas* principales:

1. Se asume la *validez* de la medida establecida con preferencia a la nueva. Una baja correlación entre la medida nueva y la establecida supone que la medida nueva es “inválida”. Pero, tal vez, la inválida sea la antigua (la tradicionalmente aceptada) o la futura, y no la nueva.
2. Muchos *conceptos* en las ciencias sociales carecen de mediciones (generalmente aceptadas) que puedan emplearse para la comprobación de nuevas mediciones. La dificultad es mayor, cuanto más abstractos sean los conceptos a medir.

En la delimitación de la *validez de constructo* fue igualmente esencial la aportación pionera de L. Cronbach y P. E. Meehl de 1955 (“Construct validity in psychological tests”. *Psychological Bulletin*, 52: 281-302), que insistieron en la adecuación de la medición de los *conceptos teóricos*. Lo bien que éstos han quedado medidos por los *indicadores* elegidos. Para su consecución es primordial la delimitación clara y precisa de los conceptos teóricos a medir. También, proceder a una *operacionalización múltiple* por dos razones principales:

1. Utilizar varias medidas para un mismo concepto contribuye a abarcar las distintas dimensiones que incluye.
2. Proporciona una mayor precisión y validez de la medición, cuando coinciden los resultados de mediciones diferentes.

Pero esto último exige que los diferentes *indicadores* que se empleen realmente midan el mismo *concepto*. Asimismo es importante la fundamentación teórica del

concepto. De acuerdo con Heath y Martin (1997: 81), “si un concepto no es parte de una teoría bien desarrollada, entonces las oportunidades para comprobar la *validez de constructo* se reducirán”.

En su defensa de la *operacionalización múltiple* como estrategia para aumentar la validez de *constructo*, Campbell y Fiske (1959) diferencian la validez *convergente* de la *divergente*:

- *Validez convergente*: cuando diferentes mediciones del mismo constructo “convergen” (correlacionan entre sí).
- *Validez divergente o discriminante*: el opuesto de la “convergente”. Los indicadores de un constructo han de mostrar bajos niveles de correlación con los correspondientes al constructo opuesto. Quiere decir que, si dos constructos son muy diferentes, las mediciones respectivas de cada uno no deberían mostrar relación entre sí.

Por ejemplo, ítems aplicados para medir xenofobia deberían dar resultados “divergentes” con aquellos seleccionados para medir xenofilia. En caso contrario, no serían indicadores “válidos”.

A ambos tipos de validez contribuye la *operacionalización múltiple*: a mayor diversidad de indicadores, mayor garantía de *validez* de los resultados. De ahí la reiterada recomendación de:

“Si no existe ningún acuerdo claro en cómo medir un concepto, médelo de varias formas diferentes. Si el concepto tiene diversas dimensiones, médelas todas. Y, por encima de todo, tienes que saber que el concepto no tiene ningún significado distinto del que le dimos” (Babbie, 1992: 134).

El análisis *multivariable* es de gran ayuda tanto en la vertiente exploratoria de descubrir las diferentes dimensiones e indicadores que agrupan (caso del análisis *factorial exploratorio* de *factor común* y de *componentes principales*), como confirmatoria (análisis *factorial confirmatorio* y el análisis *discriminante*, por ejemplo). De ellos ha de resultar la eliminación de aquellos indicadores no incluidos en alguna dimensión latente o *dominio de contenido* del *concepto* a medir.

Ya en 1983 Bohrnstedt defendía la práctica del análisis *factorial confirmatorio* como la forma más apropiada de evaluar los distintos *dominios de contenido* de un *concepto*. Por su parte, el análisis *discriminante* ha mostrado gran utilidad en la evaluación llevada a cabo por Cea D’Ancona (2002b, 2009) de los indicadores de *racismo* y *xenofobia* como ilustra el ejemplo a continuación:

EJEMPLO DE USO CONJUNTO DEL ANÁLISIS FACTORIAL Y DISCRIMINANTE

La aplicación conjunta de las técnicas analíticas *factorial* de *componentes principales* y *discriminante* fue primeramente desarrollado por M.^a Á. Cea D’Ancona en 2002 (“La medición de las actitudes ante la inmigración: evaluación de los indicadores tradicionales de racismo”. *Revista española de investigaciones sociológicas*, 99: 87-111), aunque prosigue en investigaciones posteriores.

Partiendo de las diferentes conceptualizaciones del racismo (*aversivo, simbólico, moderno, biológico, cultural, de clase*) y de las dimensiones e indicadores comúnmente presentes en su medición mediante encuesta, se evalúa la idoneidad de los presentes en la última encuesta *ex profeso* de actitudes ante la inmigración que el CIS (Centro de Investigaciones Sociológicas) realiza en junio de 1996. Con los 41 indicadores incluidos en la encuesta se realiza un análisis *factorial* de *componentes principales*. Tenía por objetivo descubrir si comparten alguna estructura latente: *dimensiones* o variables latentes que aglutinen indicadores bastante correlacionados entre sí y poco con los correspondientes a las otras *dimensiones* de racismo.

El cuadro 2.14 extracta las diez *dimensiones* que resultan del análisis tras la rotación *varimax* (ésta permite maximizar la varianza de los *coeficientes factoriales* cuadrados en cada *componente principal*, obteniendo *dimensiones* relevantes no correlacionadas entre sí). Los *componentes* figuran de acuerdo a su orden de aparición, acompañados de la etiqueta o rótulo dado por la autora para su denominación. Se ha intentado que el nombre dado a la *dimensión* latente refleje el significado de las variables que en ella más ponderan y que sintetizen su contenido “común”. Las variables aparecen acompañadas de sus *coeficientes factoriales*, cuya magnitud señala lo que caracteriza a cada componente. Su valor expresa la correlación de cada *indicador* (o variable observada) con cada *componente*. Su signo, la dirección de dicha relación: positiva o negativa.

CUADRO 2.14. Composición de los componentes principales tras la rotación *varimax*

1 DERECHOS SOCIALES		2 DERECHOS DE CIUDADANÍA	
Educación pública	0,824	Votar en elecciones municipales	0,879
Vivienda digna	0,770	Votar en elecciones generales	0,877
Asistencia sanitaria	0,747	Afiliarse a partidos políticos	0,780
Practicar su religión	0,702	Obtener la nacionalidad	0,491
Trabajo en igualdad	0,672		
Constituir asociaciones	0,669		
Traer a su familia	0,666		
Cobrar el paro	0,463		
Mantener sus costumbres	0,286		

[.../...]

CUADRO 2.14. (Continuación)

3 SOCIABILIDAD CON MARROQUÍES		4 POLÍTICA INMIGRATORIA	
Vecino marroquí	0,822	Valoración inmigración	0,614
Compañero de trabajo marroquí	0,788	Supresión fronteras UE	0,606
Casar con marroquí	0,708	Refugiado político	0,519
Colegio con inmigrantes	0,595	Control inmigración	0,509
Simpatía hacia marroquíes	-0,492	Regularizar inmigrantes	0,455
		Leyes inmigración	0,452
		Facilitar entrada a inmigrantes	0,429
		Número de inmigrantes	-0,368
5 IMAGEN TÓPICA NEGATIVA DEL INMIGRANTE		6 DISCRIMINACIÓN ÉTNICA O DE LA ALTERIDAD	
Bajan salarios	0,751	Condiciones de vida del inmigrante	-0,699
Quitán trabajo	0,750	Dificultades de los inmigrantes	0,582
Aumentan delincuencia	0,586	Trabajos no cualificados	0,536
		Ampliación de derechos	0,456
7 PARTIDO RACISTA		8 ACEPTACIÓN DEL INMIGRANTE	
Aceptación partido racista	0,766	Amistad con inmigrantes	0,668
Partido racista	0,446	Aumentará número de inmigrantes	-0,583
		Cupo de inmigrantes	-0,274
9 VIOLENCIA CONTRA INMIGRANTES		10 TEMOR AL ASENTAMIENTO DE INMIGRANTES	
Problema agresiones inmigrantes	0,817	Intención de quedarse	0,652
		Trato con desconfianza	-0,593

La efectividad de las 10 *dimensiones* se comprueba mediante un análisis *discriminante*. Ello exige el resumen de la aportación de los distintos *indicadores* que componen cada *dimensión* o componente principal en un *índice*. Para lo cual se calculan las *puntuaciones factoriales*, con el propósito de que en la representación de cada *dimensión* latente intervengan todas las variables empíricas con *coeficientes factoriales* estadísticamente significativos ($\geq 0,30$). Para su obtención se aplica el método de *Anderson-Rubin*, que propicia la obtención de *puntuaciones factoriales* no correlacionadas.

El análisis *discriminante* se realiza tomando como variable dependiente los tres grupos obtenidos del análisis de *conglomerados K-medias*, en consonancia con la actitud declarada ante la inmigración; y como independientes, las diez *puntuaciones factoriales*. Del análisis *discriminante* resulta que, en consonancia con las nuevas propuestas de *racismo*, las preguntas relativas a *derechos sociales* y de *ciudadanía* son las de mayor utilidad en la medición del *racismo* y la *xenofobia*. Le siguen los *ítems* relativos a *política migratoria* que son, a su vez, los que recogen mayor variabilidad en las respuestas de los encues-

tados. Por el contrario, los relativos a *distancia social*, presentes en la generalidad de las encuestas de actitudes y prejuicios étnicos raciales, se sitúan en el cuarto lugar en efectividad. De las diez *dimensiones* latentes sólo las seis primeras mostraron relevancia estadísticamente significativa en la medición de *racismo*. El orden lo marca el *índice de potencialidad compuesto*, que se deriva del análisis *discriminante*: *derechos sociales* (0,464), *derechos de ciudadanía* (0,044), *política inmigratoria* (0,038), *sociabilidad con marroquíes* (0,017), *discriminación étnica y de la alteridad* (0,013) e *imagen tópica negativa de la inmigración* (0,012). Para mayor información véase el susodicho artículo.

Por último, y en referencia a la *validez de constructo*, Henerson y colaboradores (1987) destacan la conveniencia de los llamados estudios de *grupo de criterio*. El instrumento de medición (escala, cuestionario...) se pasa a un grupo de personas de las que se sabe, con un cierto grado de certeza, que poseen “abundancia” o “deficiencia” del constructo que se mide. *Por ejemplo*, que en anteriores mediciones hayan alcanzado niveles elevados o bajos de racismo. En el nuevo instrumento de medida deberían alcanzar niveles similares a los antes registrados. De esta forma puede comprobarse su *validez* e identificar *indicadores* o *ítems* inadecuados para su medición.

2.5.2. Comprobación de la fiabilidad

La *fiabilidad* se refiere a “la capacidad de obtener resultados consistentes en mediciones sucesivas del mismo fenómeno” (Jacobs, 1994: 363). Se fija en la “consistencia”, en la obtención de iguales resultados utilizando el mismo instrumento de medida, con independencia del momento en que se aplique (*Por ejemplo*: pesarse en una báscula de manera repetida sin comer, beber, cambiarse de ropa... La báscula es “fiable”, si no varía el peso). Pero, que la medida resulte “consistente” –término propuesto por Smith y Stephenson (1979)– no conlleva que sea “válida”. Una medida “fiable” no siempre es “válida”, aunque una medida “válida” generalmente es también “fiable”. La *validez* es garantía de *fiabilidad*, pero no a la inversa: diferentes mediciones con un mismo instrumento pueden alcanzar los mismos resultados y ser inválidos por falta de veracidad o correspondencia con la realidad. En consecuencia sería un error demostrar *validez* a partir de la *fiabilidad*. Ambos son componentes de *calidad* diferentes y complementarios.

La forma común de comprobar la *fiabilidad* es aplicar el mismo instrumento de medición en diferentes momentos para averiguar si se alcanzan los mismos resultados. En caso negativo, deberán analizarse las causas. Existen además otros procedimientos para comprobar la *fiabilidad* que se resumieron en el cuadro 2.13, y que se relacionan con los dos conceptos con los que se identifica la *fiabilidad*: “estabilidad” y “consistencia”.

A) Método test-retest

El procedimiento tradicional y más sencillo para comprobar la *fiabilidad*. Consiste en administrar el mismo instrumento de medida a las mismas personas, en las mismas condiciones, en dos períodos de tiempo diferentes. La segunda administración ha de ser idéntica a la primera. Lo único que varía es el *tiempo*, que ha de escogerse de forma que no afecte a los resultados. Se recomienda que el tiempo entre el *test* y el *retest* sea aproximadamente de un mes, para evitar que la *memoria* o el recuerdo de la primera medición afecten a la respuesta en la segunda medición. Un período de tiempo más amplio tampoco es aconsejable. Aunque repercutiría positivamente en la reducción del efecto de la memoria, aumentaría la probabilidad de que, en ese lapso de tiempo, se produjesen cambios en la persona que alterasen la actitud, opinión o conducta que se mida. Sus respuestas en el *retest* diferirían y no se deberían a que el instrumento de medida sea “infiabile”. La *estabilidad* se exige tanto en el proceso de medición como en el fenómeno que se analiza.

La *fiabilidad* se cuantifica mediante el cálculo de los *coeficientes de correlación* de las respuestas en las dos mediciones. Un coeficiente elevado ($\geq 0,80$) implica que el instrumento de medida es “estable” y “fiable. Un valor igual a “1,0” expresa *fiabilidad* perfecta, mientras que un coeficiente próximo a “0,0” *fiabilidad* nula. Esto último sucede cuando las dos mediciones son totalmente diferentes.

$$r_{y_{i1}, y_{i2}} = \frac{\text{cov}(y_{i1}, y_{i2})}{\sqrt{\text{var}(y_{i1}) \text{var}(y_{i2})}}$$

Donde “ y_{i1} ” es la puntuación obtenida en la primera medición; “ y_{i2} ”, en la segunda.

El cálculo del *coeficiente de correlación* exige que las variables sean continuas. En caso contrario, se recomienda su previa transformación en variables *ficticias* (mediante la codificación binaria en 0 y 1 de sus atributos, como se explica en Cea D’Ancona, 2002a). Para variables categóricas puede, no obstante, aplicarse lo que Biemer y Lyberg (2003) llaman *índice de inconsistencia*, lo contrario al índice de *fiabilidad*. Es decir, “ $1 - r$ ”. Se define como la razón de la varianza de error de medición en las observaciones a la varianza total. Por lo que expresa la proporción de la varianza total que es varianza de *error de medición*.

$$I = \frac{g}{p_1(1-p_2) + p_2(1-p_1)}$$

Siendo “ g ” la proporción de casos en los que no ha coincidido la respuesta en la primera y en la segunda medición.

EJEMPLO DE CÁLCULO DEL ÍNDICE DE INCONSISTENCIA

El *índice de inconsistencia* se obtiene de una tabla de clasificación donde se cruzan las repuestas obtenidas en la primera medición con la segunda. El proceso sería como el descrito en el cuadro 2.15, con las respuestas correspondientes al ítem sobre la regularización de los inmigrantes con dos opciones de respuesta: a favor y en contra.

CUADRO 2.15

Respuesta en la segunda medición	Respuesta en la primera medición		
	A favor de la regularización	En contra de la regularización	Total
A favor de la regularización	590	37	627
En contra	23	370	393
Total	613	407	1.020

$$I = \frac{g}{p_1(1-p_2) + p_2(1-p_1)} = \frac{0,0588}{0,231 + 0,245} = 0,123$$

- Proporción de casos en los que no hay coincidencia en la respuesta en ambas mediciones:

$$g = \frac{37 + 23}{1.020} = 0,0588$$

- Proporción de casos a favor de la regularización de inmigrantes en la primera medición:

$$p_1 = \frac{613}{1.020} = 0,601$$

- Proporción de casos a favor de la regularización en la segunda medición:

$$p_2 = \frac{627}{1.020} = 0,615$$

- $p_1(1-p_2) = 0,231$ y $p_2(1-p_1) = 0,245$

Como el *índice de inconsistencia* es igual a 0,123, significa que su equivalente en *fiabilidad* sería 0,877 [(1 - 0,123) = 0,877]. Lo que implica su calificación de “fiable”.

De la tabla de clasificación puede asimismo obtenerse la proporción de casos cuya clasificación en la segunda medición concuerda con la habida en la primera (como es habitual en los análisis de *regresión logística* y *discriminante*): el *porcentaje de aciertos* en la clasificación.

$$\text{Proporción de "aciertos"} = \frac{590 + 370}{1.020} = 0,9412$$

O, lo que es igual, $(1 - g) = (1 - 0,0588) = 0,9412$

El 94,12% de los sujetos reentrevistados coincidieron en la respuesta que dieron en la primera medición. Un porcentaje de acuerdo que se estima bastante elevado.

Pero, antes de adjudicar “fiabilidad” o “infiabilidad” al instrumento de medida (a partir de coeficientes de *correlación* o del *índice de inconsistencia*), habría que eliminar explicaciones alternativas de la coincidencia o no coincidencia de las respuestas porque:

- Un *coeficiente de correlación bajo* puede deberse a cambios en el individuo, entre la primera y la segunda medición.

Por ejemplo, que en ese período de tiempo se haya establecido relación con inmigrantes y las respuestas recabadas en la segunda medición sean más favorables a la inmigración que las registradas en la primera. También, puede suceder a la inversa: una mala experiencia, personal o ajena (como actos de violencia protagonizados por inmigrantes), puede propiciar un mayor rechazo expreso a la inmigración en la segunda medición. Cambios en las respuestas no atribuibles al instrumento de medida, sino a cambios en la persona que responde a las preguntas.

A mayor distancia temporal entre el *test* y el *retest*, más probable es la presencia de esta fuente de invalidez.

- *Acortar en exceso el período de tiempo* entre ambas mediciones puede provocar sobreestimación de la *fiabilidad* por el efecto de la *memoria*: acordarse de las respuestas dadas en la primera medición y responder, en el *retest*, de la misma forma para mostrar “consistencia” en la respuesta. Habitual cuanto menor sea el período de tiempo entre *test* y *retest*.

Además puede darse el efecto contrario: que la *memoria* lleve a dar respuestas diferentes para parecer flexibles, de mente abierta, especialmente en ítems actitudinales (Groves, 1989). Aunque esto último parece menos factible (Carmines y Zeller, 1979; Bohrnsted, 1983; De Vaus, 1990; Krosnick y Fabrigar, 1997). Es más probable que la *memoria* contribuya a la *consistencia* que a la *inconsistencia* de las respuestas.

- Por último, hay que considerar el efecto debido a *reactividad*: la primera medición puede provocar sensibilización en la persona hacia el tema sobre el que se pregunta y que ésta modifique su respuesta en el *retest* (que dé respuestas más meditadas, fruto de la reflexión sobre lo antes preguntado y respondido). Asimismo incide el estar ya familiarizado con el instrumento de medida, al no variar éste en el *retest*.

A estas matizaciones se suma la dificultad de aplicar, de forma reiterada, un mismo instrumento de medida a una misma población. Factores *internos* (cansancio, pérdida de interés) y *externos* (cambios de domicilio, de ocupación u otra modificación en sus circunstancias vitales) que impidan contactar de nuevo con la persona para el *retest*. Su repercusión en los resultados dependerá de la disparidad que exista entre aquellos que continúan participando y los que han abandonado el estudio. Un problema que adquiere un mayor alcance conforme se amplía el tiempo entre el *test* y el *retest*. En general, este método de comprobación de la *fiabilidad* se desaconseja siempre que se analicen fenómenos inestables o más expuestos a cambios. En esos casos habría que optar por otros métodos de comprobación de la *fiabilidad*.

B) Método alternativo

Para evitar los efectos de la *memoria* en el *retest* se aplica un método alternativo, que mida el mismo concepto de distinta forma.

Por ejemplo, dividir aleatoriamente 60 ítems de *xenofobia*: 30 se incluyen en la escala aplicada en el *test* y los 30 restantes para el *retest*. Ha de procurarse la similaridad y distribución aleatoria de los ítems en ambas escalas para que puedan compararse las respuestas (a través de coeficientes de *correlación* entre ambas mediciones).

Precisamente es ésta su mayor dificultad: crear ítems que sean realmente comparables. En palabras de Bohrnstedt (1983: 80), “pocas ‘formas paralelas’ de medición son realmente paralelas”. A ello se suma otra deficiencia importante: no logra resolver, de forma clara, la distinción entre el cambio “verdadero” de la “infiabilidad” de la medición (Carmines y Zeller, 1979), al igual que el método *test-retest*.

C) Método de las dos mitades

Las dos comprobaciones se realizan en sujetos diferentes y al mismo tiempo (de manera simultánea). Para ello se divide la serie de ítems en dos mitades, que pueden ser equilibradas o aleatorias. *Aleatorias*, cuando la disposición de la serie

de preguntas en las dos mitades es al azar. *Equilibradas*, si se busca la paridad entre ellas, procurando (mediante casación) que ítems “iguales” estén en ambas mitades.

Al igual que el anterior, su práctica exige garantizar que ambas mitades sean homogéneas (para evitar que las correlaciones entre ellas se deban a diferencias de contenido real). Deben suscitar la misma capacidad o actitud en los sujetos. Su éxito dependerá de que exista una serie amplia de ítems y que puedan repartirse equitativamente en dos mitades (que el constructo esté igualmente representado en ambas mitades).

Por ejemplo, los 60 ítems de *xenofobia* primero se aplican a un grupo de personas. Después se divide los ítems al azar en dos grupos de 30, y se pasan a dos grupos diferentes de personas (pero que compartan iguales características), y al mismo tiempo. Se calculan los coeficientes de *correlación*. Una correlación elevada ($\geq 0,70$, preferiblemente, $0,80$) supone “consistencia interna” y fiabilidad.

Este método pretende resolver los problemas detectados en los dos métodos anteriores, separando las consideraciones de *fiabilidad* de los efectos de la memoria, el aprendizaje o la propia evolución del individuo. No obstante, se sugiere la necesidad de aplicar un ajuste estadístico que corrija el tamaño pequeño de las dos series de ítems que se comparan (véase Lyons Morris *et al.*, 1987; o Henerson *et al.*, 1987).

D) Método de la consistencia interna alfa de Cronbach

Éste es el método para comprobar la *fiabilidad* que más se aplica debido a su inclusión en los paquetes estadísticos de uso común. Como expresa su nombre, indaga en la “consistencia interna”, en la homogeneidad de los indicadores que se aplican para medir un *concepto*. Los indicadores que corresponden a una misma dimensión latente han de estar intercorrelacionados entre sí y diferir de los indicadores de las otras dimensiones latentes.

La *consistencia interna* se identifica con el coeficiente *alfa* propuesto por Lee J. Cronbach en 1951 (en “Coefficient alpha and the internal structure of tests”. *Psychometrika*, 16: 297-334). Aunque G. F. Kuder y M. W. Richard fueron los primeros en recomendar una medida de equivalencia que utilizase todas las covarianzas entre los ítems, en 1937 (“The theory of the estimation of test reliability”. *Psychometrika*, 2: 135-138).

El *coeficiente alfa* se calcula a partir de la *matriz de varianzas-covarianzas* de la serie de ítems que componen una escala de medida. A diferencia de la matriz de correlación, en la matriz de varianzas-covarianzas las variables no están estandarizadas. La *diagonal* de la matriz la forman las varianzas de cada ítem. El resto de la matriz son las *covarianzas* entre pares de ítems. El *coeficiente alfa* informa de la va-

rianza “común” o compartida entre cada par de *ítems*. Su valor se obtiene de la fórmula siguiente:

$$\alpha = \left[\frac{k}{k-1} \right] \left[1 - \frac{\sum_{i=1}^k S_i^2}{S_t^2} \right]$$

Donde “ k ” es el número de *ítems*; S_i^2 es la varianza del ítem i y S_t^2 es la varianza de la suma de todos los ítems.

El *coeficiente alfa* va desde “0,0” (infiabilidad) hasta “1,0” (fiabilidad perfecta). Expresa la proporción de la varianza total de la serie de ítems incluidos en la escala de medida que se debe a la variable latente, siendo en consecuencia “común”. Por esta razón, el *coeficiente alfa* “se corresponde estrechamente con la definición clásica de fiabilidad como la proporción de varianza en una escala que es atribuible a la puntuación verdadera de la variable latente” (De Vellis, 1991: 41). En general se recomienda un valor $\geq 0,80$ para que el instrumento de medida se considere “fiable”; al menos ha de superar el valor de “0,70”.

Al igual que la *validez*, existen diferentes fuentes de *error* que puedan afectar a la *fiabilidad*: errores de *medición* debidos al diseño del instrumento de medida (ítems incluidos, su redacción y ubicación en la escala, en el cuestionario...), su aplicación (autocumplimentación, mediante entrevista: actuación y características de los entrevistadores...), la actitud de las personas a las que se aplica (estado de ánimo, cansancio, interés, imagen que quieran proyectar, conocimiento del tema, efecto de la memoria), y cómo se analizó la información (codificación, grabación, análisis estadístico). Aunque tampoco hay que descuidar los errores de *no observación*, debidos al perfil sociodemográfico de los participantes en la investigación y los excluidos del estudio. Lo cierto es que la *fiabilidad* “perfecta” es difícil de alcanzar. En palabras de Carmines y Zeller (1979: 11, 12):

“La medición de cualquier fenómeno siempre contiene una cierta cantidad de error causal (...) Porque las mediciones repetidas nunca igualan exactamente unas a otras, la infiabilidad siempre está presente, al menos a una extensión limitada”.

Para conseguir un coeficiente de *fiabilidad* elevado cuatro son las actuaciones comúnmente recomendadas:

- a) La *muestra* seleccionada (para comprobar la *fiabilidad*) ha de ser *amplia* y *heterogénea* respecto al objeto que se mide. Ello favorece la obtención de coeficientes de correlación elevados. También conviene que la *muestra* se adecue a la población a la que finalmente se aplicará el instrumento de medida.
- b) *Operacionalización múltiple*: cuantos más indicadores se incluyan, y del mayor número de dimensiones o “dominios” del constructo, mejor. Se impone

la elección de indicadores relevantes para la medición del concepto, eliminando todo aquel que no esté relacionado con el objeto de estudio.

Por ejemplo, si se realiza un *análisis factorial*, se eliminará todo indicador que presente una baja *comunalidad* (o valor R^2 : el coeficiente de correlación múltiple cuadrado obtenido de regresionar la variable con el resto de las incluidas en el análisis); es decir, una baja proporción de varianza compartida con los demás indicadores.

La eliminación de ítems de escasa relevancia suele repercutir en incrementos en el valor del coeficiente *alfa*. En general, circunscribir el análisis sólo a aquellos indicadores más relacionados con el objeto de estudio contribuye a la obtención de estimaciones muestrales de los parámetros poblacionales más estables, porque disminuyen los *errores típicos de estimación*.

- c) *Aumentar la varianza de las variables* incrementando las opciones de respuesta. Cuanto mayor sea su número, más precisa será la medición.
- d) *Estandarizar el proceso de recogida de información*. Además reduce la varianza debida al entrevistador.

“La mejor forma de demostrar que tu trabajo ha sido mínimamente contaminado por inconsistencias de “instrumentos humanos” es utilizar más de una persona para hacer al menos una muestra de tus entrevistas u observaciones. Si personas diferentes informan bastante la misma cosa, eso es evidencia de consistencia” (Henerson *et al.*, 1987: 149).

En general reducir cualquier *error* que merme la *calidad* de la medición.

EJEMPLO DE COMPROBACIÓN DE VALIDEZ Y FIABILIDAD DE INSTRUMENTOS DE MEDIDA

Comprobación de la *validez* y *fiabilidad* de dos instrumentos de medida que han mostrado ser eficaces a la hora de predecir el divorcio o separación matrimonial: la *Escala de Inestabilidad Matrimonial* – EIM de A. Booth y J. Edwards (1983: “Measuring marital instability”. *Journal of Marriage and the Family*, 45: 387-393) y el *Índice de Satisfacción Matrimonial* – ISM de W. W. Hudson (1992: *The WALMYR Assessment Scales Scoring Manual*. Tempe: WALMYR Publishing). En ambos casos realizado por Joseba Iraurgi, Mireya Sanz y Ana Martínez-Pampliega en 2009 (“Adaptación y estudio psicométrico de dos instrumentos de pareja: índice de satisfacción matrimonial y escala de inestabilidad matrimonial”. *Revista IIPSI*, vol. 12, n.º 2: 177-192).

El objetivo del estudio era comprobar la adaptación cultural de ambos instrumentos de medida analizando cinco tipos de equivalencia entre la escala original y la escala adaptada:

- Equivalencia *conceptual*: qué dominios son importantes para el concepto y las relaciones entre ellos.
- *De ítems*: examinar los ítems empleados para esos dominios, que podrían variar entre culturas.
- *Semántica*: asegurarse de que la traducción logra equivalencia semántica de los ítems.
- *Operacional*: que los métodos de medida empleados sean apropiados para la cultura en cuestión.
- *De medida*: examinar los resultados del proceso en términos del comportamiento del instrumento.

La *muestra* la formaban 183 mujeres. Todas tenían hijos de edades comprendidas entre 8 y 18 años. En su mayoría fueron familias con uno (29,5%, $n = 54$) o dos hijos (56,8%, $n = 104$). Previamente se sometieron los instrumentos a un proceso de traducción y adaptación conceptual a su contexto cultural. Se evaluó la equivalencia conceptual y la claridad del enunciado de los ítems hasta llegar a una versión adaptada.

Para el estudio de la *fiabilidad* de los instrumentos se analizó la *consistencia interna* de cada uno de ellos y de sus dimensiones teóricas y empíricas a través del *coeficiente alfa* (α) de Cronbach. Éste calibra si todos los ítems de una escala miden una misma dimensión (coeficientes elevados, de 0,90 y 0,97 para el ISM y EIM, respectivamente). Para analizar la *validez de constructo* se realizó un análisis *factorial exploratorio de componentes principales* (con rotación varimax); después un análisis factorial *confirmatorio* para la verificación de las dimensiones extraídas. La *validez concurrente* se comprueba a través del análisis de asociación del ISM y EIM con medidas de otros constructos con las que se esperaba que tuviesen relaciones significativas basándose en determinados modelos teóricos y la investigación previa realizada. Se calcularon las *correlaciones de Pearson* (r) entre los instrumentos bajo estudio y los cuestionarios ASPA (*Cuestionario de aserción en la pareja*, de Carrasco), VR-SLH (*Listado de Síntomas* de Hopkins, que analiza cinco factores: somatización, depresión, ansiedad, dificultades cognitivas y sensibilidad interpersonal), y la *escala de satisfacción familiar* de Olson y Wilson (evalúa el grado de satisfacción experimentado con aspectos relacionados con la cohesión y adaptabilidad familiar). Se obtuvieron coeficientes de correlación por encima de 0,40. Todo lo cual llevó a la conclusión de que ambas escalas cumplen sólidas propiedades psicométricas y muestran gran adecuación para la predicción de la ruptura de la pareja.

Lecturas complementarias

- Abad, F. J.; Olea, J.; Ponsoda, V. y García, C. (2011): *Medición en ciencias sociales y de la salud*. Síntesis. Madrid.
- Carmines, E. G. y Zeller, R. A. (1979): *Reliability and validity assessment*. Sage. Beverly Hills.

- Corbetta, P. (2010): *Metodología y técnicas de investigación social*. McGraw-Hill. Madrid.
- DeVellis, R. F. (1991): *Scale development. Theory and applications*. Sage. Newbury Park.
- González Blasco, P. (1994): “Medir en las ciencias sociales”, en García Ferrando, M. *et al.*, *El análisis de la realidad social. Métodos y técnicas de investigación*. Alianza. Madrid: 275-334.
- Henerson, M. E.; Morris, L. L. y Fitz-Gibbon, C. T. (1987): *How to measure attitudes*. Sage. Newbury Park, California.
- Neuman, W. L. (2000): *Social research methods: qualitative and quantitative approaches*. Allyn and Bacon. Boston, MA.

Ejercicios propuestos

- Operacionalice los conceptos *discriminación* y *felicidad*: definiciones teóricas y operativas, dimensiones e indicadores.
- Elabore ítems a incluir en una escala de *xenofobia*. Adóptelos a los diferentes formatos de escala para la medición de actitudes.
- ¿En qué difiere la validez *concurrente* de la *convergente*? Ilustre los diferentes tipos de *validez* con un ejemplo.
- Destaque las ventajas e inconvenientes de las distintas formas de comprobar la *fiabilidad*.
- Calcule un *índice* que resuma la información incluida en la siguiente tabla:

Pobl. 65 años y más con hijos (% horizontal)	¿Cómo calificaría la relación con sus hijos?					Índice
	Muy buena	Buena	Regular	Mala	Muy mala	
Varones	17	33	31	14	5	
Mujeres	35	34	21	8	2	
– 75 años	31	35	28	4	2	
+ 75 años	23	25	30	16	6	
Total	27	31	29	9	4	

3

LA LÓGICA EXPERIMENTAL EN LA INVESTIGACIÓN SOCIAL

La *experimentación* es la *estrategia* de investigación social que más se fundamenta en los principios del positivismo. Aunque es más ampliamente utilizada en psicología, su aplicación se extiende a otras áreas de conocimiento. Neuman (2000: 222) destaca su aportación en “educación, justicia, periodismo, marketing, enfermería, ciencia política, trabajo social, y sociología”. Despunta como la *estrategia* que más se adecua a la comprobación de relaciones causales. Principalmente, por el mayor control que introduce de las explicaciones alternativas a una relación causa-efecto analizada.

En las páginas que siguen se va a ir desgranando la lógica experimental, sus características esenciales y la pluralidad de formas que puede adoptar en la investigación social, además de sus potencialidades y límites.

3.1. Fundamentos de la experimentación

Originaria de las ciencias naturales, la introducción de la *experimentación* en la investigación social comienza en el terreno de la psicología, de la mano de Wilhelm Maximilian Wundt (1832-1920). Un fisiólogo, filósofo estructuralista y psicólogo alemán que creó el primer laboratorio experimental en 1879 en Leipzig. En él se realizaron experimentos sobre sensaciones y percepciones (memoria de imágenes, ilusiones visuales...). Pero la expansión del *método experimental* en la investigación social llegaría de la mano del:

- *Conductismo* (o *behaviorismo*): escuela psicológica creada por John Broadus Watson (1878-1958) en 1913, con la publicación de su artículo “Psychology as the behaviorist views” (en *Psychological Review*, 20, 158-177), y posterior-

mente desarrollada por Burrhus Frederic Skinner (1904-1990) y su análisis de la “conducta operante” (la capacidad del organismo para organizar los estímulos y la conducta en términos de medios-fines). Defendió la medición de la conducta observable (no la mente) como el objeto de estudio de la psicología y del *método experimental* para desarrollar rigurosos contrastes empíricos de hipótesis.

- *Cuantificación* (o medición de los fenómenos sociales con números), que también se desarrolló en el período de 1900 a 1940: se reconceptualizaron constructos sociales para que pudiesen ser cuantificados, mientras que otros (como la voluntad, la conciencia, el espíritu...) fueron descartados de la investigación empírica. Despuntan la medición de la capacidad mental mediante los tests de inteligencia (IQ), originalmente desarrollados Alfred Binet (1857-1911), psicólogo y pedagogo francés que diseñó el *test de predicción del rendimiento escolar* (junto con Théodore Simon), que sirvió de base para los tests sucesivos de inteligencia. De acuerdo con Neuman (2000: 224), entre los años 1921 y 1936 más de 5.000 artículos se publicaron sobre tests de inteligencia. Período que también coincide con el desarrollo de las técnicas escalares y de medición, y los avances de la estadística en la investigación social (como se irá viendo en los capítulos sucesivos).

A éstas se suman otras contribuciones relevantes como la realizada por el estadístico Ronald A. Fisher a la fundamentación estadística de la *experimentación*. Destaca su aportación a la asignación “aleatoria” a los grupos experimentales, y otros principios de la experimentación, tanto en su libro *Statistical Methods for Research Workers* (1925) (considerado uno de los libros sobre métodos estadísticos más relevantes), como en *The Design of Experiments* (1935). Este último ilustra cómo diseñar experimentos desde el punto de vista estadístico. Además ha de resaltar su contribución básica a la técnica analítica principal en la experimentación: el análisis de la *varianza* (descrito en el capítulo 8).

También destacables son las aportaciones pioneras de Donald T. Campbell y Julian C. Stanley (1963) a la configuración de los diseños *experimentales* y *cuasiexperimentales* en la investigación social y los criterios de *validez* (algunos ya expuestos en el capítulo 1 y otros tratados en el apartado 3.4).

3.2. Características esenciales de la experimentación

La *experimentación* es una estrategia de investigación que se basa en el *control* e *intervención* (por parte del investigador) de la “realidad” que analiza. Consiste en la comprobación de los efectos de una o varias variables (características de un sujeto u objeto), que se *manipulan* (variables *independientes* o predictoras - X), en la ocurrencia del fenómeno cuya variabilidad se analiza (variable *dependiente* - Y): actitudes, conductas, sentimientos, creencias, que cambian en respuesta al *tratamiento* (o

variable *independiente*). La medición de la variable *dependiente* puede hacerse mediante observación, entrevista, escalas, cuestionarios, respuestas fisiológicas (como palpitaciones, gritos, llantos...). Además de *manipular* las variables que presumiblemente son su “causa” (a partir de una o varias *hipótesis* a comprobar), se *controlan* variables extrañas (que puedan mediar en la relación entre las variables X e Y) y se aleatorizan las restantes. El objetivo es excluir cualquier explicación alternativa a las variables manipuladas en la causación del fenómeno que se analiza.

El estudio de la *causalidad* no sólo exige el cumplimiento de los tres criterios fundamentales del concepto de *causalidad* enunciados por David Hume (1711-1776), en su *Tratado sobre la Naturaleza Humana* (1739-1740):

- *Contigüidad* entre la *causa* y el *efecto* (ambas variables han de covariar: variar conjuntamente).
- *Precedencia temporal* de la *causa* sobre el *efecto* (la variable independiente ha de acontecer antes que la dependiente).
- *Conjunción* constante entre la *causa* y el *efecto* (siempre que se presenta la “causa” sucede el “efecto” y, a la inversa, cuando no aparece la “causa”, tampoco se observa el “efecto”).

A éstos se suma el requisito de la exclusión de posibles explicaciones alternativas a la relación *causa-efecto* observada, que ya fuera apuntado por John Stuart Mill en 1843, en su obra *Sistema de lógica*. En ella se especifican pilares principales de la *causalidad* que aún se comparten: determinismo, universalidad, contigüidad y prioridad temporal. En la medida en que se cumpla este último requisito, la indagación alcanzará *validez interna* (como se viera en el capítulo 1). En la investigación mediante *encuesta* la eliminación de explicaciones alternativas se alcanza mediante *control a posteriori* (una vez recogida la información), durante el proceso de análisis. En la *experimentación*, el *control* es *a priori*, antes de proceder a la recogida de datos. La información se obtiene mediante una o varias *técnicas* de investigación (observación sistemática, cuestionario, entrevista). A través de ellas se recaba información que permita medir los cambios que la *manipulación experimental* provoca en el fenómeno que se estudia. Tres son sus características diferenciales:

- a) *Manipulación experimental*: el investigador crea una situación (artificial) que le permita comprobar la influencia causal de una o más variables (*independientes*) en otra variable (la *dependiente*). La variable *independiente* (cuyo efecto trata de medirse) adquiere la forma de *tratamiento* o *estímulo experimental* (el nombre proviene de la medicina). El número de *tratamientos* varía en función de cuántos valores o categorías incluya la variable *independiente* (dando forma a diferentes diseños experimentales, que se exponen en el apartado 3.3).
- b) *Control del investigador*: crucial en la investigación experimental. El investigador ha de “controlar” cualquier factor que pueda incidir en la relación

causa-efecto que se analiza: aislar los efectos del *tratamiento* y eliminar las explicaciones alternativas. La *manipulación experimental* está condicionada a las posibilidades de controlar (o eliminar) la influencia de cualquier variable que pueda incidir en la relación causal entre dos variables (y que no interese al estudio). Sin este requisito, no cabe hablar de influencia causal.

“Cuando se está interesado en comprobar si una manipulación determinada en algún aspecto del entorno tiene algún efecto, se necesita controlar, o mantener constante, otros aspectos del entorno” (Jones, 1985: 286).

Aunque no es probable que puedan “controlarse” todos los factores que afectan a una relación *causal*, han de mantenerse constantes (controlar) todos aquellos que se prevean (por indagaciones previas, por juicio...) que puedan afectar.

- c) *Aleatorización*: un componente esencial en todo experimento. Está presente en dos momentos fundamentales del proceso experimental: cuando se asigna cada sujeto a alguno de los grupos constituidos; y cuando se asigna cada grupo a una de las condiciones experimentales (Morales *et al.*, 1981).

En un *experimento* habrá, al menos, dos *grupos*: el grupo *experimental* (al que se aplica el tratamiento, cuyo efecto trata de medirse), y el grupo de *control* (al que se le priva de dicho tratamiento, o se le asigna un tratamiento *placebo*). Mediante la formación de este segundo grupo, se intenta “controlar” el efecto de factores ajenos al estímulo, o *tratamiento experimental*, que puedan afectar a los resultados del estudio (características internas y externas a los individuos incluidos en la investigación). Para ello ha de procurarse la total *equivalencia* inicial de ambos grupos: que los integrantes de los grupos (*experimental* y de *control*) compartan las mismas características. Sólo han de diferir en el nivel de *tratamiento* que posteriormente se les aplica. En la medida en que esto se consiga, se podrá separar los efectos (en la variable *dependiente*) provocados por la manipulación de la variable *independiente*.

La *equivalencia* entre los grupos *experimental* y de *control* puede alcanzarse mediante:

- *Igualación* (o casación) razonada de los sujetos en aquellas variables que se prevea puedan afectar al valor de la variable *dependiente*. Se ha de eliminar (o neutralizar) su influencia, asegurando su presencia, en la misma magnitud, en los diferentes grupos que componen el experimento. El problema principal de este procedimiento de *control* (de variables ajenas al estudio) es que no hay plena certeza de haber alcanzado la “igualación” de los grupos *experimental* y de *control*, y en variables que puedan afectar a los resultados del estudio. Los casos pueden diferir de muchas formas y el investigador no siempre sabe (*a priori*) cuáles podrían afectar a la relación *causa-efecto* analizada.

- La *aleatorización* o asignación aleatoria de los sujetos a los diferentes grupos favorece su equivalencia inicial. La distribución de los sujetos no responde a criterios del investigador, sino al “azar”. Aunque puedan existir leves diferencias entre los grupos, la *aleatorización* garantiza su equivalencia “aproximada” en un número elevado de características (no sólo en aquellas que se prevean y controlen mediante la *igualación*). Además, la *aleatorización* proporciona a todos los sujetos la misma probabilidad de formar parte de uno u otro grupo. Dicha probabilidad puede ser matemáticamente determinada (como se verá en el capítulo 7).

Por ejemplo, se quiere comprobar el efecto del método de enseñanza en el nivel de conocimientos adquirido por los alumnos. La equiparación inicial de los grupos de *tratamiento* y de *control* mediante la *igualación* (o casación) supone la distribución de los individuos en los grupos según los valores que presenten en una serie de variables que se prevé pueden mediar en la relación causa-efecto que se estudia: la calificación media del expediente académico, su cociente de inteligencia, el grado de motivación al éxito, las horas dedicadas al estudio... Se procura su reparto equitativo (en igual proporción) en los diferentes grupos. Pero, por muchas características que se abarquen, siempre habrá alguna variable cuyo efecto perturbador no se controle, al no haberse previsto (caso de características relativas al ámbito donde estudia, o a la situación económica, familiar o afectiva del alumno). En cambio la distribución de los sujetos mediante un procedimiento aleatorio (*simple*, *sistemático* u otro explicado en el capítulo 7) propicia que esas variables no previstas también queden “controladas” o neutralizadas.

La precisa equivalencia inicial de los grupos puede comprobarse antes de aplicar el *tratamiento* (mediante un *cuestionario* u otro instrumento que permita calibrar si los distintos grupos parten de iguales características iniciales). De esta forma podrá eliminarse cualquier explicación alternativa de los resultados del estudio que puedan atribuirse a la composición diferencial de los grupos *experimental* y de *control*; en caso contrario, no habrá plena certeza.

Además de las características personales y sociales de los sujetos observados, habrá que “controlar” las características ambientales o contextuales que puedan asimismo propiciar explicaciones alternativas a la relación *causa-efecto* que se analiza. En el *ejemplo* anterior, habrá que igualar las características de las aulas donde se imparten métodos de enseñanza diferentes (iluminación, audición, franja horaria...). Cuantos más factores contextuales se controlen, mayor seguridad habrá en los efectos de los tratamientos introducidos. Es el minucioso *control* “*a priori*”, que exige la práctica de un *experimento*, lo que contribuye a la *artificialidad* atribuida como característica definidora de la investigación experimental. Si bien, el *control* es requisito básico para la comprobación de relaciones causales y lo que convierte a la *experimentación* en la estrategia idónea para el análisis de la *causalidad*.

Una vez concluido el *experimento*, se procede a la *comparación* de los resultados en los grupos diferenciados para comprobar los efectos de los *tratamientos*. La

comparación de los resultados grupales en la variable *dependiente* puede realizarse mediante un análisis de *varianza*, u otra técnica analítica. Si bien, en la investigación *experimental* es muy usual la aplicación del análisis de la *varianza* para medir las diferencias de las *medias* grupales en la variable *dependiente* y su significatividad estadística. Los *tests de significatividad* (el estadístico *t*, en el análisis de la *varianza simple*; y *F*, en el análisis *múltiple* de la *varianza*) indican si las diferencias observadas entre las *medias* grupales son o no estadísticamente significativas: si se deben o no a diferencias en la composición de los grupos. Cuando carecen de significatividad estadística, no puede afirmarse que el *tratamiento* experimental haya tenido efecto. Por el contrario, cuando las diferencias entre los grupos son estadísticamente significativas, no cabe atribuirlos a divergencias en la composición de los grupos. Afecta el *tamaño* de la *muestra* que se analiza y el procedimiento seguido para distribuir los individuos en los grupos. De acuerdo con Saxe y Fine (1981), si se hace mediante *aleatorización*, descendiende la probabilidad de que los resultados del experimento realmente se deban a características particulares del entorno, del experimentador, o de la población expuesta al *estímulo experimental*.

EJEMPLO DE EXPERIMENTO

Para ilustrar la lógica experimental, se reseña un experimento clásico que ha adquirido una amplia repercusión tanto en psicología como en sociología; en concreto, en las áreas de sociología industrial, psicología social del trabajo, psiquiatría industrial y antropología del trabajo: los *experimentos de Hawthorne*. Fueron dirigidos por George Elton Mayo (1880-1949), un sociólogo y psicólogo industrial especializado en teoría de las organizaciones y las relaciones humanas, interesado en el análisis de los efectos psicológicos de las condiciones físicas del trabajo en la producción. Contó con la colaboración de Leta Stetter Stouffer, el ingeniero de la *Western Electric* William Dickson y personal de la Universidad de Harvard. Los resultados de los experimentos fueron publicados en 1933 (*The social problems of an industrial civilization*).

La Academia Nacional de la Ciencia de EE UU desde 1924 patrocina estudios en empresas para analizar los factores que afectan a la productividad de los trabajadores. En 1927 encarga a Elton Mayo que dirija un estudio en la *Western Electric Company* (una fábrica de equipos y componentes telefónicos), en el barrio de Hawthorne de Chicago. Una empresa que desarrollaba una política de personal dirigida al bienestar de los obreros, pagaba salarios satisfactorios y brindaba buenas condiciones de trabajo. En su departamento de montaje de relés de teléfono trabajaban jóvenes empleadas que realizaban tareas simples y repetitivas que exigían gran rapidez. En aquella época, una empleada montaba cinco relés cada seis minutos. La empresa no estaba interesada en aumentar la producción, sino en conocer mejor a sus empleadas y las condiciones ambientales que afectaban a su actividad. El estudio comprende cuatro fases:

En la *primera fase* se seleccionaron dos grupos de trabajadoras que realizaban exactamente la misma labor, en condiciones idénticas. En el grupo *experimental* se

manipuló la iluminación reduciéndola durante períodos sucesivos. En el grupo de *control* (con el que se comparaban los efectos, y que trabajaba en una zona separada), la iluminación se mantuvo constante. El objetivo del experimento era medir el efecto de la iluminación en el rendimiento de los trabajadores. Se partía de la hipótesis de que la producción aumenta cuando los operarios trabajan con mejor iluminación.

En ambos grupos el desempeño se elevaba de manera constante, a pesar de que la iluminación en el grupo *experimental* se hizo tan reducida que los trabajadores se quejaron de que apenas podían ver. Fue así que bajo este mínimo nivel de iluminación, el desempeño de este grupo inició una reducción sensible. Ante esta evidencia empírica, los investigadores concluyeron que ante variaciones “normales” de iluminación (la escasísima luz fue un caso extremo), la productividad no variaba en igual proporción. En ambos grupos (el *experimental* y el de *control*) habían influido otros factores no identificados y distintos a la iluminación. Factores de carácter *psicológico*, pues los operarios reaccionaban a la experiencia de acuerdo a sus suposiciones personales: se creían en la obligación de producir más, cuando la intensidad de la luz aumentaba, y viceversa. Este hecho se comprobó cambiando lámparas por otras de igual intensidad, pero haciéndoles creer que eran de mayor fuerza. Así, se pudo verificar la existencia de un nivel de rendimiento proporcional a la intensidad de la luz sobre la cual las operarias “suponían” que trabajaban. Factores psicológicos parecían ser determinantes. Asimismo se conjeturó que ambos grupos podrían haber estado en contacto, aun cuando estaban distantes uno de otro, y ello podría haber propiciado la competencia entre ellos.

Los sorprendentes resultados de la primera fase impulsaron un *segundo experimento* dirigido a analizar los efectos de la mejora de las condiciones laborales. Se seleccionaron 6 mujeres jóvenes con un nivel medio de experiencia (cinco montaban relés, mientras que la sexta suministraba las piezas necesarias para mantener un trabajo continuo y eficiente). Éstas forman el grupo *experimental*, dedicadas a ensamblar relés eléctricos en un cuarto separado del resto del departamento (grupo de *control*). La separación era una pared de madera. Tanto el equipo como la mesa de trabajo eran idénticos, pero tenían un plano inclinado con un contador de piezas individual que indicaba la producción de cada operaria. El indicador más fácilmente medible era el nivel de producción, y por ello pasó a ser el parámetro de comparación entre ambos grupos. El *experimental* estaba sujeto a cambios en las condiciones de trabajo; el de *control* no: trabajaba siempre bajo las mismas condiciones. El grupo *experimental* tenía un supervisor común al grupo de *control*, pero además contaba con un observador que permanecía en la sala, ordenaba el trabajo y se encargaba de mantener el espíritu de cooperación y ayuda mutua de las operarias. El observador contó con la colaboración de algunos asistentes, a medida que se hacía más complejo el experimento. A ambos se les informó de los objetivos del estudio: determinar el efecto de ciertos cambios en las condiciones de trabajo (período de descanso, reducción del horario de trabajo, comidas, incentivos...); se les informaba de los resultados, y se sometían a su aprobación las modificaciones que fueran a introducirse. Se insistía en que trabajasen con normalidad y pusieran voluntad en el trabajo. La investigación con el grupo *experimental* se dividió en 12 períodos experimentales para poder averiguar cuáles eran las condiciones más satisfactorias de rendimiento.

Previamente, al inicio formal del experimento, los investigadores previeron las posibles reacciones negativas del personal participante en él. Intentando reducir las posibles resistencias, los investigadores modificaron el estilo usual de supervisión de modo que no existiera un supervisor oficial de la empresa. La idea era que los trabajadores se de-

sempañaran bajo la observación de los propios investigadores. Adicionalmente, se otorgaron privilegios “especiales” a los trabajadores, tales como abandonar sus puestos de trabajo sin permiso, así como recibir atención personal de los experimentadores e inclusive de los propios funcionarios de la empresa.

En el primer período se registra la producción de cada trabajadora (duró 2 semanas); en el segundo, se aisló al grupo *experimental* en la sala de pruebas y se mantuvieron normales las condiciones y el horario de trabajo (duró 5 semanas y permitió comprobar el efecto del cambio de ubicación del trabajo); en el tercero, se modificó el sistema de pagos (duró 8 semanas y aumentó la producción); en el cuarto, se introdujeron cinco minutos de descanso a mitad de la mañana y otros cinco a mitad de la tarde (aumentó la producción); en el quinto, los intervalos de descanso aumentaron a diez minutos (la producción aumentó); en el sexto, se dieron tres descansos de cinco minutos por la mañana y otros tres por la tarde (la producción no aumentó y hubo quejas de que se rompía el ritmo de trabajo); en el séptimo, se volvió a los intervalos de diez minutos, uno por la mañana y otro por la tarde, y durante uno de ellos se sirvió un refrigerio ligero (aumentó la producción); en el octavo, con las mismas condiciones del período anterior, el grupo *experimental* comenzó a trabajar hasta las 16:30 y no hasta las 17:00 como el grupo de *control* (hubo un acentuado aumento de la producción); en el noveno, la jornada laboral del grupo *experimental* concluía a las 16:00 (la producción permaneció estable); en el décimo, el grupo *experimental* volvió a trabajar hasta las 17:00, como en el séptimo período (la producción aumentó considerablemente); en el undécimo, el grupo *experimental* no trabajaba el sábado (la producción continuaba subiendo); en el duodécimo, se volvió a las condiciones del tercer período. Este último período duró 12 semanas e inesperadamente se observó que la producción alcanzó niveles jamás logrados (3.000 unidades semanales por trabajadora en el grupo *experimental*).

Los investigadores de la Universidad de Harvard, que participaban como evaluadores del experimento, concluyeron que la productividad se elevó debido a la modificación en la forma en que se llevó a cabo la supervisión. También se consideró que los cambios físicos como el plan de incentivos, las comidas, la jornada de trabajo, la paga..., no tuvieron un impacto decisivo en los resultados. Se comprobó que incluso desmejorando las condiciones de trabajo, la producción seguía aumentando. Los investigadores tipificaron lo que después se llamaría “*el efecto Hawthorne*”: la posibilidad de que los individuos que se separan para un estudio puedan mejorar su desempeño simplemente por la atención adicional que reciben de los investigadores, y no debido a cualquier factor específico que se probara en el estudio. Las jóvenes manifestaron que les gustaba trabajar en una sala de pruebas, que la supervisión era menos rígida y que les permitía trabajar con más libertad y menor ansiedad. A pesar de que en la sala de pruebas existía más supervisión que en el departamento (donde había apenas un supervisor para un número mayor de trabajadoras), la característica y el objetivo de la supervisión eran diferentes y las jóvenes lo sabían. En especial, sentían que participaban en un experimento interesante y que debían producir resultados que, aunque no los conociesen bien, redundarían en beneficio de las demás compañeras. Además, en el grupo *experimental* las jóvenes entablaron relaciones de amistad (que se extendían más allá del ambiente laboral), se preocupaban por las otras, acelerando su producción cuando alguna compañera demostraba cansancio. Constituían un grupo que desarrolló liderazgo y objetivos comunes. Después de que salieran dos jóvenes del grupo original, una de las sustitutas se volvió espontáneamente líder, ayudando a sus compañeras a alcanzar el objetivo común de aumentar continuamente el ritmo de producción, a pesar de que constantemente se les pedía que trabajaran con normalidad.

En la *tercera fase*, los investigadores (preocupados por la diferencia de actitudes observadas entre los grupos *experimental* y de *control*) fueron abandonando poco a poco su interés en las variables físicas. Gradualmente centraron su atención en variables *psicológicas* y en las *relaciones humanas* en el trabajo. En septiembre de 1928 se inició el programa de *entrevistas* para conocer mejor sus actitudes, sentimientos, opiniones sobre su trabajo y el tratamiento que recibían. La empresa contaba con 40.000 trabajadores y entre 1928 y 1930 se realizaron 21.126 entrevistas. En 1931 el sistema de entrevista adoptó la forma de *entrevista no dirigida* (que permitía a los entrevistados hablar libremente, sin que el entrevistador les condujera o estableciera pautas previas). Los investigadores observaron que, además de la organización formal de la empresa, había una informal que resultaba de las relaciones entre los miembros de la empresa, y que se basaba en sentimientos comunes y expectativas. Con objeto de profundizar en esta organización informal, se desarrolló una cuarta fase de la investigación.

Para la *cuarta fase*, que duró de noviembre de 1931 a mayo de 1932, se seleccionó un grupo *experimental* de 9 operadores, 9 soldados y 2 inspectores, todos de la línea de montaje de terminales para estaciones telefónicas. Todos ellos pasaron a trabajar a una sala especial, equipada con idénticas condiciones de trabajo que las del departamento. Había un entrevistador dentro de la sala y un entrevistador fuera de ésta. Esporádicamente, conversaban con los operarios. El objetivo era observar el comportamiento informal del grupo en condiciones variables de trabajo (su organización informal y relación con la estructura formal de organización). Se comprobó que los trabajadores presentaban cierta uniformidad de sentimientos y solidaridad grupal: consideraba delator al miembro que perjudicaba a algún compañero y presionaba a los más rápidos, a través de penalizaciones simbólicas, para “estabilizar” su producción.

Los *experimentos de Hawthorne* delinearon los principios básicos de la *escuela de las relaciones humanas*. Mostraron que el nivel de producción no está determinado por la capacidad física o fisiológica de los empleados, como afirmaban los partidarios del *taylorismo* y del *foyolismo*, sino por las normas sociales y por las expectativas que éstas involucran. El comportamiento del individuo se apoya totalmente en el grupo. Las personas no son motivadas tanto por los incentivos salariales como por la necesidad de “reconocimiento”, de “aprobación social”, y de “participación” en las actividades de los grupos sociales donde conviven. De ahí la transición del “homo oeconomicus” al “hombre social”. Los aspectos formales de la organización no son tan importantes como los informales. La empresa pasa a considerarse una organización social compuesta por grupos sociales informales, cuya estructura no siempre coincide con la formal (los fines y estructura definidos por la empresa).

Por último, destacar los *efectos* que pudieron repercutir en los resultados de los *experimentos de Hawthorne*:

- “*Efecto Hawthorne*”: una forma de reactividad psicológica por la que los sujetos de un experimento muestran una modificación en algún aspecto de su conducta como consecuencia del hecho de saber que están siendo estudiados, y no como respuesta a alguna *manipulación* del estudio experimental.
- Efecto de la *mortalidad experimental*: dos de las trabajadoras fueron sustituidas durante el transcurso del experimento. Los datos de productividad pudieron verse afectados por la alteración de la *muestra* inicial.
- Efecto del *aprendizaje* que las trabajadoras pudieron acumular a lo largo de cinco años, y que podría explicar la mejora de su rendimiento.

- La *retroalimentación*: las trabajadoras del grupo *experimental*, a diferencia del resto de sus compañeras, tenían acceso a la información sobre su rendimiento individual.
- El hecho de formar parte de un grupo reducido pudo favorecer que las trabajadoras asignasen un valor más alto a su propio rendimiento individual, manteniendo unos niveles más elevados de motivación (entendían que su papel en el grupo no se reducía a una pequeña aportación dentro de un grupo más numeroso).

3.3. Pluralidad de diseños experimentales

Una vez concretados los objetivos e hipótesis a comprobar, el investigador ha de decidir qué diseño experimental aplicar dentro de las limitaciones concretas del estudio. El cuadro 3.1 sintetiza la variedad de *diseños experimentales*. Éstos se configuran a partir de la aportación pionera y fundamental de Campbell y Stanley (1970).

CUADRO 3.1. Tipología de diseños experimentales

<i>Experimentos intersujetos</i>	De dos grupos aleatorios	<ul style="list-style-type: none"> – Diseño solamente después – Diseño antes y después – Diseño Solomon
	De más de dos grupos	
	Diseños factoriales	<ul style="list-style-type: none"> – Con dos variables independientes – Con más de dos variables independientes
<i>Experimentos intrasujetos</i>	<ul style="list-style-type: none"> – Dos tratamientos experimentales, un solo grupo – Varios tratamientos experimentales, un solo grupo – Diseños factoriales intrasujetos 	

Los experimentos *intersujetos* son los que predominan en la investigación social. Como se deduce de su nombre, suponen la formación de distintos grupos de sujetos, a los que se aplican *tratamientos* diferentes y cuyos efectos se comprueban. Comprende las tres variedades siguientes:

- Experimentos de dos grupos aleatorios*: la modalidad experimental más básica. Consiste en la formación aleatoria de sólo dos grupos: el grupo *experimental* (al que se aplica el *tratamiento* cuyo efecto se quiere comprobar) y el grupo de *control* (utilizado de referente para “controlar” los efectos de variables *perturbadoras* en la relación *causa-efecto* que se analiza). Los sujetos se distribuyen,

de forma aleatoria, en cada uno de los grupos. Estos experimentos pueden adoptar tres formas diferentes en función de cuántas veces se mida el valor de la variable dependiente (cuya variabilidad quiere analizarse):

- *Diseño solamente después o posttest (X O)*: la variable *dependiente* sólo se mide (O = observación de la variable *dependiente*), en los dos grupos constituidos, una vez aplicado el *tratamiento* experimental (X = *tratamiento*, variable *independiente*). La inexistencia de medición previa al *tratamiento* se convierte en su principal limitación: no hay plena certeza de la equivalencia inicial de los dos grupos (en el valor de la variable *dependiente*). Se confía en que ésta quede garantizada por la asignación aleatoria de los sujetos a los grupos.
- *Diseño antes y después (O X O)*: el diseño experimental clásico. Incluye la asignación aleatoria al grupo *experimental* y de *control*, además de la medición *pre* y *post* *tratamiento*. La variable *dependiente* se mide *antes* y *después* de la aplicación del *tratamiento*, y en ambos grupos: *experimental* (el que recibe el *tratamiento*) y de *control* (con el que se compara para la medición de los efectos). De este modo se garantiza que ambos grupos sean equivalentes antes de la aplicación del *tratamiento*. Por lo que las diferencias observadas en el valor de la variable *dependiente*, entre la primera y la segunda medición pueden atribuirse al efecto del *tratamiento* (o variable *independiente*: X).

La *ventaja* principal de la *premedición* (*pretest*) es que permite conocer el valor de la variable *dependiente* antes y después del *tratamiento* (o intervención experimental), y en ambos grupos. El *inconveniente*, que los sujetos pueden percibir la finalidad del estudio y modificar su conducta (no ajustándose a su proceder natural). De la agudeza e ingenio del investigador depende que logre enmascarar el objetivo del estudio para evitar el efecto no deseado que su conocimiento pudiera tener en la respuesta de los sujetos al *tratamiento* experimental. Por ejemplo, si la *premedición* (*pretest*) se hace a través de un *cuestionario*, incluir una amplia variedad de preguntas, que logren despistar al encuestado sobre el propósito del estudio.

- *Diseño Solomon (O X O / X O)*: combina las dos modalidades anteriores. Su denominación le viene del investigador que lo ideó: Richard L. Solomon (1949: “An extension of control group design. *Psychological Bulletin*, 46: 137-150). Su propósito era controlar el efecto de la medición *pretratamiento*.

Consiste en la formación aleatoria, no de dos grupos, sino de cuatro: dos grupos *experimentales* y otros dos de *control*. A uno de los grupos *experimental* y de *control* se les mide el valor de la variable *dependiente* antes y después de la aplicación del *tratamiento*; a los otros dos grupos (uno *experimental* y otro de *control*) sólo cuando el *tratamiento* ha concluido. Ello permite comprobar el efecto que la *medición* inicial puede tener en los resultados de la investigación. Otra *ventaja* importante es la posibilidad de comprobar, de diferentes formas, una misma *hipótesis* de investiga-

ción (Morales *et al.*, 1981). El *inconveniente* principal es la dificultad de formar cuatro grupos de sujetos.

- b) *Experimentos de más de dos grupos*: cuando la variable *independiente* incluye más de dos valores. El número de grupos dependerá del número de categorías o valores diferenciados en la variable *independiente*. Como en los experimentos de dos grupos, en éstos también se analiza la relación causal entre una única variable *independiente* y una sola variable *dependiente*.
- c) *Diseños factoriales*: analizan al mismo tiempo las relaciones causales entre más de una variable *independiente* y una sola variable *dependiente*. Cada variable *independiente* incluye un variado número de valores. El número de *grupos de tratamiento* estará en función del número de variables *independientes* consideradas, y de cuántos valores comprendan. Habrá un grupo por cada valor de cada variable *independiente*. Los sujetos se distribuirán de manera aleatoria entre los diferentes grupos como en los demás diseños experimentales.

Por ejemplo, si se estudiase la influencia en la atención del alumno en clase (variable *dependiente*) de dos variables *independientes* (duración de la clase y tipo de docencia), y cada una con tres opciones posibles, se estaría ante un *diseño factorial 3×3*: habría seis *grupos de tratamiento*. Cada uno resulta de la conjugación de una categoría de cada variable como puede observarse en el siguiente cuadro:

CUADRO 3.2. Diseño factorial 3×3

Tipo de docencia (A)	Duración de la clase (B)		
	Una hora	Dos horas	Tres horas
Teórica	A ₁ B ₁	A ₁ B ₂	A ₁ B ₃
Teórico-práctica	A ₂ B ₁	A ₂ B ₂	A ₂ B ₃
Práctica	A ₃ B ₁	A ₃ B ₂	A ₃ B ₃

Si se añadiese una tercera variable *independiente* (turno de docencia) con dos opciones (mañana y tarde), se estaría ante un *diseño factorial 3×3×2* = 18 grupos de tratamiento. Y así consecutivamente.

Los diseños *factoriales* permiten comprobar varias hipótesis causales en una misma investigación. Los efectos de las variables *independientes* pueden, asimismo, comprobarse por separado y de manera conjunta. Ello ayuda al análisis de efectos recíprocos de dos o más variables *independientes*. El *inconveniente* principal es la complejidad que el experimento adquiere a medida que aumenta el número de variables *independientes* y los valores que éstas incluyen.

EJEMPLO DE EXPERIMENTO FACTORIAL

Experimento sobre los *efectos sociocognitivos de las noticias* sobre inmigración, realizado por Juan-José Igartua, Carlos Muñiz, José Otero, Lifen Cheng y José Gómez-Isla (2008: “Recepción e impacto sociocognitivo de las noticias sobre inmigración”. *Revista de Psicología Social*, 23 (1), 3-16).

Tomando como referencia la investigación sobre el efecto de los encuadres noticiosos, dos fueron los *objetivos* específicos del estudio:

1. Analizar en qué medida los encuadres más *problematizadores* de la inmigración (que vinculan inmigración con delincuencia), frente a encuadres que enfatizan aspectos más positivos (contribución económica), influyen en la percepción de la inmigración como problema social y condiciona las opiniones y las creencias sobre las consecuencias de la inmigración.
2. Evaluar la influencia del origen geográfico de los inmigrantes protagonistas de la información noticiosa en el procesamiento y los efectos provocados por los encuadres noticiosos.

Se plantearon dos *hipótesis*:

- H1: Los sujetos expuestos a la noticia con encuadre de contribución económica generarán más respuestas cognitivas positivas hacia la inmigración, más respuestas cognitivas centradas en el encuadre de consecuencias económicas y menos centradas en el encuadre de conflicto que los sujetos expuestos a la noticia con un encuadre que asocia inmigración con delincuencia. Este resultado será menos acusado cuando los inmigrantes protagonistas de la noticia sean de origen marroquí.
- H2: El encuadre que alude a la contribución económica de los inmigrantes, frente al que vincula inmigración con delincuencia, inducirá una menor percepción de la inmigración como problema, una opinión más favorable hacia la inmigración y un mayor grado de acuerdo con creencias que relacionan la inmigración con el progreso económico del país. Este resultado será menos acusado cuando los inmigrantes protagonistas de la noticia sean de origen marroquí.

Se utilizó un *diseño factorial 2 x 2*, siendo las variables *independientes* el tipo de encuadre enfatizado en la noticia y el tipo de colectivo inmigrante aludido de manera incidental. Se utilizaron dos tipos de encuadres noticiosos: los inmigrantes como delincuentes y la contribución económica de los inmigrantes. Además, en la mitad de las noticias se aludía a inmigrantes marroquíes; en la otra mitad, a latinoamericanos.

En el estudio participaron 186 estudiantes de la Universidad de Salamanca de nacionalidad española. El 75,8% de los sujetos eran mujeres; la media de edad fue de 20,95 años (rango: 18-38 años). En cada aula, los sujetos fueron aleatorizados a las cuatro condiciones experimentales. Se obtuvieron entre 45 y 48 cuestionarios en cada condición experimental. El tiempo total dedicado a la lectura de la noticia y cumplimentación del cuestionario fue de 20-25 minutos. El *trabajo de campo* se realizó en tres días durante el mes de octubre de 2005. En la portada del cuestionario se entregaba información sobre

los objetivos del estudio. En la segunda página se incluía la noticia experimental que en las cuatro condiciones presentaba el mismo titular: “*España alcanza los 4 millones de extranjeros, el 9% de la población*”. En las páginas siguientes se incluían las variables dependientes, de chequeo de la manipulación y de control.

Las noticias utilizadas se construyeron tomando como referencia investigaciones previas de *análisis de contenido*, y a partir de noticias publicadas sobre el tema en el momento del estudio. En las cuatro noticias construidas (que contenían entre 592 y 620 palabras) el titular, entradilla (*lead*), primer párrafo y cierre o conclusión (cuarto párrafo) era el mismo. Se refería al incremento de la inmigración en España, aportando datos del Instituto Nacional de Estadística sobre la situación actual y las proyecciones a futuro.

La manipulación de los encuadres noticiosos estuvo determinada por la alusión a diferentes tipos de consecuencias asociadas a la inmigración. En una de las versiones se aludía a las consecuencias negativas de la inmigración para el país. Se vinculaba el texto con informaciones sobre el incremento de la delincuencia atribuido a la inmigración. Se informaba de un incidente de carácter delictivo ocurrido en Madrid, protagonizado por un inmigrante, en el que había muerto un ciudadano autóctono. En otra versión se destacaban las consecuencias positivas de la inmigración para el país en el terreno económico. Se resaltaban efectos como el incremento de los ingresos en la Seguridad Social y la revitalización de los pequeños comercios en zonas donde residían inmigrantes, como Madrid.

En tres momentos diferentes a lo largo de la noticia se incluía información sobre el origen geográfico o nacional de los inmigrantes. En la presentación de los datos estadísticos sobre la presencia de inmigrantes en España se enfatizaban diferentes informaciones. En una de las versiones se afirmaba que el primer colectivo afincado en España era el compuesto por marroquíes, mientras que en otra versión se indicaba que el primer colectivo era el compuesto por inmigrantes latinoamericanos. A lo largo de la noticia se aludía en otras dos ocasiones a inmigrantes de procedencia marroquí o latinoamericana en el relato de diferentes hechos. En la versión con un encuadre de conflicto se describía el mismo acto delictivo cometido, en uno de los casos, por un ciudadano marroquí, y en otro, por un ciudadano latinoamericano.

A los sujetos se les pidió: “escribe todos los pensamientos, ideas o reflexiones que te haya suscitado la lectura de la noticia, es decir, aquellas impresiones que te surgieron durante la lectura de la noticia”. Tomando como unidad de análisis cada una de las ideas escritas, dos analistas evaluaron los criterios:

- Si en la respuesta cognitiva estaba presente el encuadre de conflicto, mediante comentarios sobre la relación inmigración-delincuencia, la inseguridad ciudadana, el aumento de la delincuencia, referencia a delitos cometidos por inmigrantes, detenciones de inmigrantes, conflictos en los barrios por la inseguridad (1 = sí, 0 = no).
- Si en la respuesta cognitiva estaba presente el *encuadre de consecuencias económicas*, mediante comentarios sobre la mejora de la economía, los ingresos de la Seguridad Social, el fortalecimiento de la población activa o el incremento de la actividad comercial (1 = sí, 0 = no).
- *Polaridad de la respuesta cognitiva en relación a la inmigración*: comentario crítico o negativo (valor -1, por ejemplo, “el aumento de inmigrantes está aumentando la delincuencia”), comentario ambivalente, no evaluativo o de tipo descriptivo (valor 0, por ejemplo, “la mayoría de los inmigrantes son sudamericanos”) o comentario

favorable (valor 1, por ejemplo, “el tema de las pensiones es un buen argumento para rebatir a los que piensan que la inmigración sólo trae el caos y la delincuencia”).

Para evaluar la *fiabilidad* del proceso de *codificación*, se eligieron al azar 28 cuestionarios (el 11% de la *muestra*; lo que supuso un total de 101 respuestas cognitivas), y se codificaron de manera independiente los criterios mencionados. Se calculó el porcentaje de acuerdo entre los dos jueces, arrojando los siguientes valores: alusión al encuadre de conflicto (97%), alusión al encuadre de consecuencias económicas positivas (90%) y polaridad (80%). A partir de la información de las respuestas cognitivas individuales se generó un fichero de datos agregados que fue fusionado con el resto de variables dependientes.

A los sujetos se les preguntaba por: el tema principal de la noticia, el porcentaje de inmigrantes en España según la noticia, el origen geográfico de los inmigrantes protagonistas de la noticia, así como otras informaciones de interés. Se creó un *índice* de memorización de la noticia, sumando las opciones contestadas correctamente.

- *Importancia de la inmigración como problema.* El texto de la pregunta era: “Por favor, indica según tu opinión en qué medida los siguientes asuntos son problemas importantes para el país” (siendo la *inmigración* uno de los 13 temas). Para cada asunto se indicaba el nivel de importancia (0 = nada, 10 = muy importante).
- *Opinión general hacia la inmigración.* Se preguntó: “Como sabes, todos los países desarrollados reciben inmigrantes. ¿Crees que, en términos generales, la inmigración es más bien positiva o más bien negativa para España?”. Los sujetos indicaban su opinión utilizando una escala de 11 puntos (0 = muy negativa, 10 = muy positiva).
- *Creencias sobre las consecuencias de la inmigración.* Se construyó una escala con 8 afirmaciones sobre las consecuencias de la inmigración en España. Se preguntaba por el grado de acuerdo o desacuerdo (1 = total desacuerdo, 5 = total acuerdo) con las afirmaciones: a) “Los inmigrantes desempeñan trabajos que los españoles no quieren”; b) “En España, hoy en día, se siguen necesitando trabajadores inmigrantes”; c) “El aumento de la inmigración favorece el incremento de la delincuencia”; d) “Al aceptar sueldos más bajos, los trabajadores extranjeros hacen que bajen los salarios de los españoles”; e) “Existe una estrecha relación entre inmigración e inseguridad ciudadana”; f) “Los inmigrantes quitan puestos de trabajo a los españoles”; g) “El aumento de los inmigrantes favorece la economía del país”; h) “En general, los inmigrantes están contribuyendo al desarrollo de España”.

Un análisis factorial de *componentes principales* (rotación *varimax*) extrajo dos factores. El primer factor (ítems *c*, *d*, *e* y *f*) aludía a la creencia “la inmigración favorece el incremento de la delincuencia y la competencia desleal a los trabajadores españoles”; el segundo (ítems *a*, *b*, *g* y *h*) se refería a la creencia “la inmigración supone una contribución económica para el país”.

- *Autoposicionamiento político.* Se evaluó mediante una escala de 10 puntos, desde 1 (izquierda) hasta 10 (derecha).
- *Consumo de medios.* Se preguntó: “En términos generales, en un día normal de la semana ¿cuántas horas ves televisión? ¿Escuchas radio? ¿Lees la prensa de información general? ¿Navegas por Internet?”
- *Nivel de contacto con inmigrantes.* Se preguntaba al sujeto si mantenía, o había mantenido en el pasado (1 = sí, 0 = no), algún tipo de relación familiar, de amistad, de trabajo o de vecindad con inmigrantes. Se creó un índice de contacto personal

con inmigrantes, a partir de la suma simple de las cuatro variables dicotómicas consideradas (rango teórico de puntuaciones desde 0 hasta 4).

No se observaron diferencias estadísticamente significativas entre las cuatro condiciones experimentales en las variables sexo, edad, autopostricionamiento político y nivel de contacto con inmigrantes. Tampoco se observaron diferencias significativas en el consumo de televisión, radio, prensa e Internet, ni en el índice de memorización de la noticia. Estos resultados avalaron el procedimiento de *aleatorización* y sugirieron que la *manipulación experimental* fue efectiva.

Con respecto a las respuestas cognitivas de la relación entre inmigración y delincuencia (H_1), se observó un efecto significativo del tipo de encuadre: los sujetos expuestos a una noticia con un encuadre de conflicto elaboraron más respuestas cognitivas que vinculaban inmigración y delincuencia que los sujetos expuestos a un encuadre de consecuencias económicas. No se observó un efecto significativo asociado al origen de los inmigrantes, pero sí un efecto de interacción significativo entre ambas variables *independientes*.

Cuando la noticia enfatizaba un encuadre de consecuencias económicas positivas se generaban más respuestas cognitivas centradas en el conflicto, si los protagonistas eran marroquíes que si eran latinoamericanos

También se obtuvieron diferencias estadísticamente significativas en la percepción de la importancia de la inmigración como un problema para el país en función del tipo de encuadre (H_2). No existía un efecto significativo en función del origen de los inmigrantes, ni tampoco un efecto de interacción. Los sujetos expuestos a la noticia que vinculaba inmigración y delincuencia, frente a los sujetos expuestos a una noticia que enfatizaba las consecuencias económicas positivas, concedían una mayor importancia a la inmigración como problema para el país.

El origen de los inmigrantes no influía en la opinión general hacia la inmigración. Tampoco se observó un efecto de interacción significativo. Los sujetos expuestos a una noticia que hacía saliente la contribución económica de los inmigrantes manifestaban una opinión más positiva que los sujetos expuestos a una noticia con un enfoque de conflicto.

La creencia positiva “la inmigración supone una contribución económica para el país” estaba afectada por el tipo de encuadre: los sujetos expuestos a una noticia que hacía saliente la contribución económica de los inmigrantes mostraban un mayor acuerdo con dicha creencia que aquellos expuestos a una noticia que vinculaba inmigración con delincuencia. No se apreció un efecto significativo en función del origen de los inmigrantes. Sin embargo, se observó un efecto de interacción significativo: ante una noticia con encuadre de consecuencias positivas, y con protagonistas latinoamericanos, se mostraba un mayor acuerdo con esta creencia que si la misma noticia estaba protagonizada por marroquíes.

Los resultados pusieron de manifiesto que la manera de enfocar el tema de la inmigración generaba efectos cognitivos y actitudinales; lo que era congruente con las investigaciones previas sobre el efecto de los encuadres noticiosos. Tres *limitaciones* se destacan del estudio:

1. La noticia experimental no presentaba material gráfico de ningún tipo (en futuros estudios sería recomendable, dado que los encuadres noticiosos también se manifiestan mediante recursos visuales y pueden provocar un mayor impacto).

2. Se debería comprobar el impacto de otros tipos de encuadres noticiosos de la inmigración de carácter negativo, que tienen una alta presencia en los medios de comunicación (la percepción de avalancha, de entrada masiva y de invasión).
3. La investigación se desarrolló en Castilla y León, una comunidad española con baja densidad de población inmigrante. Se debería comprobar en qué medida el efecto de los encuadres noticiosos sobre inmigración está moderado por la densidad de la población inmigrante en la comunidad de residencia de los sujetos experimentales.

Los *experimentos intrasujetos* difieren de los *intersujetos* en que los diferentes *tratamientos* se aplican a un *único grupo* de sujetos. A las mismas personas se las somete, en momentos diferentes, a *varios tratamientos*. La finalidad es analizar su reacción diferencial a los mismos. Las disparidades que se observen en el valor de la variable *dependiente*, una vez concluido el *tratamiento*, se consideran una medida de su efecto.

Este tipo de diseño experimental comparte los *inconvenientes* comunes a cualquier análisis de una misma población de forma reiterada:

1. El *desgaste de la muestra*, con el consiguiente aumento de la “no respuesta”.
2. Los *sesgos* provocados por las mediciones anteriores en las posteriores. A la persona se le dan más oportunidades para adivinar el objetivo de la investigación, con la consiguiente posibilidad de que modifiquen su conducta en busca de consonancia con la actuación anterior u otra razón.

Como se aplican *varios tratamientos* a los mismos sujetos, con la dificultad consiguiente de separar los efectos debidos a los distintos tratamientos, el investigador ha de controlar, de manera especial, el orden de presentación de los *tratamientos*. Los *tratamientos* que pudieran sesgar más las respuestas a tratamientos posteriores deberían aplicarse al final del *experimento*. De modo que los tratamientos menos reactivos se aplicasen primero.

Además de la graduación de los *tratamientos* en función de su grado de *reactividad*, el investigador podría optar por dividir la *muestra* en submuestras (al menos dos). En cada una de ellas se aplicarían los mismos tratamientos, pero en orden inverso. De esta forma se podrían medir los efectos reactivos de los diferentes tratamientos.

Al igual que los experimentos *intersujetos*, los *intrasujetos* presentan diferentes modalidades en función del número de variables *independientes* y los valores o categorías que éstas incluyan. Se diferencia entre:

- a) *Experimentos intrasujetos de un solo grupo y dos tratamientos experimentales*: sólo existe una variable *independiente* con dos valores o categorías. A los

sujetos de la *muestra* se les mide el valor de la variable *dependiente* antes y después de los dos tratamientos. Las diferencias que se observen entre ambas mediciones se atribuyen al efecto de los valores de la variable *independiente* en la *dependiente*. Hay sólo un grupo *experimental*, que actúa, a su vez, de grupo de *control*.

- b) *De un solo grupo y varios tratamientos experimentales*: si la variable *independiente* incluye más de dos valores.
- c) *Diseños factoriales intrasujetos*: cuando se miden efectos conjuntos de varias variables *independientes*.

A la tipología anterior de *diseños experimentales* se suma la distinción básica entre experimentos de *campo* y de *laboratorio*:

- a) Los *experimentos de campo* se realizan en contextos “reales” o *naturales* (contraponiéndose a la artificialidad que caracteriza a los experimentos de *laboratorio*), con el consiguiente menor *control* de las condiciones experimentales. Pueden desarrollarse en fábricas, aulas escolares, hospitales, centros penitenciarios, parques públicos o cualquier lugar donde quiera analizarse la reacción habitual o normal ante determinados hechos o conductas. El investigador puede encontrarse con una determinada situación de manera azarosa (*por ejemplo*, un accidente de tráfico), y observar cuál es la reacción de la gente (si socorren a las víctimas, huyen, o únicamente observan impávidos). O también puede crear la situación.

Por ejemplo, poner a una persona herida tumbada en una calle comercial para, a continuación, observar (como un transeúnte más) cuál es la reacción de la gente que pasa junto a la persona herida.

En los *experimentos de campo* no se realiza una selección aleatoria de los sujetos o unidades de la *muestra*. Se observa el hacer de personas corrientes en su cotidianidad. Ello exige que las personas desconozcan que participan en un experimento. Además, el propio entorno “natural” (donde se desarrolla la investigación) dificulta el *control* experimental a priori, limitando la *validez interna* del *experimento* (o posibilidad de eliminar explicaciones alternativas a la relación *causa-efecto* que se analiza).

Por el contrario, mejora la *validez externa* (o *representatividad* de la *muestra* que se observa respecto al conjunto de la población). De ahí, la semejanza de los experimentos de *campo* con los *cuasi-experimentos*, más que con *experimentos* propiamente dichos (Cook, 1983).

- b) Los *experimentos de laboratorio* se ajustan a los principios básicos de la *experimentación*. Puede “manipularse” cualquier situación que se estime relevante para comprobar las hipótesis de investigación. Se controlan las características que han de reunir las unidades de la *muestra*, al igual que las

singularidades ambientales que pudieran influir en los resultados del *experimento*.

Fuentes de *invalidéz* de los *experimentos de laboratorio*:

- A diferencia de los experimentos de *campo*, en los de *laboratorio* los sujetos suelen conocer su participación en el experimento. Ello puede provocar un *inconveniente* importante: la modificación de su conducta; que ésta se ajuste a lo que “se espera” y no a su proceder normal.

En 1962, Martin Theodore Orne demuestra el *efecto de las características de la demanda* en los resultados de la investigación: el deseo de realizar una contribución “útil” a la investigación lleva a los sujetos a ajustar su conducta a lo que se espera de ellos (corroborando las *hipótesis* del estudio).

- Las *características del experimentador* también afectan a la *validez interna* de los experimentos de *laboratorio*. En 1966, Robert Rosenthal muestra cómo el experimentador puede convertirse en una fuente importante de *sesgo* en los experimentos de *laboratorio*. Características personales (sexo, edad, etnia) y su actuación pueden influir en los resultados de la investigación. El investigador puede proporcionar a los sujetos experimentales señales que les indiquen si su reacción al *estímulo experimental* se adecua a las expectativas de la investigación. Entre estas *señales* destacan el tono de voz y gestos (como asentir con la cabeza, girarla, cerrar los ojos, fruncir el ceño). Además están las actuaciones relacionadas con su función de observador y registrador de las respuestas a los estímulos experimentales.
- Una fuente importante de *invalidéz externa* de los experimentos de *laboratorio* es que, en su mayoría, se realizan con estudiantes universitarios (de los primeros cursos de psicología, preferiblemente). Estos estudiantes a veces se ofrecen voluntarios, otras veces se les insta a ello (por cursar estudios de sociología o de psicología), y pueden o no percibir una remuneración por su participación en el *experimento*. Ello genera problemas en la generalización de los resultados (*validez externa*), debido a que estos estudiantes no representan al conjunto de la *población*. A veces, incluso, ni a los estudiantes en general.

La mayoría de los voluntarios son varones, con una necesidad elevada de aprobación social, y suelen ser más sociables y menos convencionales que aquellos estudiantes que no se ofrecen voluntarios (Orenstein y Phillips, 1978). Las hipótesis experimentales deberían comprobarse en una amplia diversidad de sujetos (y no restringirse a personal voluntario).

- La propia *artificialidad* creada en los experimentos de *laboratorio* (para el *control* de explicaciones alternativas) le resta *validez externa*; es decir, la posibilidad de generalización de los hallazgos más allá de los confines del laboratorio. Esto provoca argumentos en contra de su utilidad, precisamente por la “artificialidad” del entorno donde éstos acontecen (Bryman, 1995).

EJEMPLO DE EXPERIMENTO DE LABORATORIO

Como ilustración de experimento de laboratorio se referencia uno de los más famosos de la Psicología Social: El *experimento de la prisión de Stanford* de Philip G. Zimbardo (1973: "The psychological power and pathology of imprisonment", en Aronson, E. y Helmreich, R., *Social Psychology*. Van Nostrand. Nueva York).

Un experimento sobre la influencia de un ambiente extremo (vida en prisión) en la conducta del hombre, en función del rol social que desarrolla (preso, guardián). Se realizó en el verano de 1971, por un equipo de investigadores (profesores y estudiantes de psicología social de la Universidad de Stanford), liderado por Zimbardo, en una prisión ficticia que construyeron en una sección de algo más de diez metros y medio de un pasillo en los sótanos del edificio de Psicología de dicha Universidad. La limitaban dos paredes que habían sido construidas al efecto. En una de ellas estaba instalada la única puerta de entrada al bloque de celdas; en la otra, una pequeña pantalla de observación.

Se convirtió un conjunto de salas de laboratorio en tres pequeñas celdas de 2,10 metros por 3,15 metros, poniendo en lugar de las puertas originales otras con barrotes de acero pintados de negro y eliminando todo mobiliario. En un ala adyacente del edificio se habilitaron varias salas para alojar a los guardias, un dormitorio para el guarda jefe y el superintendente, y una sala para entrevistas. Detrás de la pantalla de observación había un equipo de vídeo y espacio suficiente para varios observadores. Se hizo *observación* directa y continuada de todas las interacciones. A ello se sumó la aplicación de *cuestionarios*, *escalas de autoinforme* y *entrevistas*.

El estudio fue subvencionado por la Armada de EE UU, interesada en buscar una explicación a los conflictos en su sistema de prisiones y en el del Cuerpo de Marines. Zimbardo y su equipo intentaron probar la hipótesis de que los guardias de prisiones y los convictos se autoseleccionaban, a partir de una cierta disposición que explicaría los abusos cometidos frecuentemente en las cárceles. Un investigador asistente (David Jaffe) sería el "alcaide" y Zimbardo el "superintendente". Después, Zimbardo reconocería que, desde un punto de vista metodológico, la acumulación en su persona de la condición de investigador principal y del rol de superintendente de la prisión fue un grave error.

Los participantes fueron reclutados a través de anuncios publicados en el periódico local, donde se ofrecía una paga de 15 dólares diarios por participar en la "simulación de una prisión". Además se señalaba que los participantes debían firmar un contrato aceptando una serie de condiciones (como que el experimento duraría entre cinco días y dos semanas, que sólo podrían abandonar la prisión por motivos de salud, o que la conducta de unos y otros sería captada mediante imágenes que podrían ser emitidas con fines científicos). Según el anuncio, el *experimento* cumplía un doble objetivo: desarrollar normas de control del comportamiento y observar la conducta de personas en principio comparables (proceden de una misma población), pero a las que arbitrariamente se les asignan roles diferentes.

De las 70 personas que respondieron al anuncio, Zimbardo y su equipo seleccionaron a los 24 que estimaron más saludables y estables psicológicamente. Los participantes eran predominantemente blancos, jóvenes y de clase media; estudiantes norteamericanos y canadienses de dicha universidad. El grupo de 24 jóvenes fue dividido aleatoriamente en dos mitades: los "prisioneros" y los "guardias". Más tarde los prisioneros dirían que los guardias habían sido elegidos por tener la complexión física más robusta,

aunque en realidad se les asignó el papel lanzando una moneda al aire y no había diferencias objetivas de estatura o complexión entre los dos grupos.

Zimbardo estableció varias condiciones específicas que esperaba que provocaran la desorientación, la despersonalización y la desindividualización. Los guardias recibieron porras y uniformes caqui de inspiración militar, que habían escogido ellos mismos en un almacén militar. También se les proporcionaron gafas de espejo para impedir el contacto visual (Zimbardo dijo que tomó la idea de la película *Cool hand Luke - La leyenda del indomable*). A diferencia de los prisioneros, los guardias trabajarían en turnos y volverían a casa durante las horas libres, aunque durante el experimento muchos se prestaron voluntarios para hacer horas extra sin paga adicional.

Los prisioneros debían vestir sólo batas de muselina (sin ropa interior) y sandalias con tacones de goma, que Zimbardo escogió para forzarles a adoptar “posturas corporales no familiares” y contribuir a su incomodidad para provocar la desorientación. Se les llamaría por números en lugar de por sus nombres. Estos números estaban cosidos a sus uniformes. Además debían llevar medias de nailon en la cabeza para simular que tenían las cabezas rapadas, a semejanza de los reclutas en entrenamiento, y una pequeña cadena alrededor de sus tobillos como “recordatorio constante” de su encarcelamiento y opresión.

El día anterior al experimento, los guardias asistieron a una breve reunión de orientación, pero no se les proporcionó ninguna formación específica para el desempeño de sus funciones carcelarias. Ninguna regla explícita aparte de la prohibición de ejercer la violencia física. Se les dijo que era su responsabilidad mantener el orden en la prisión simulada, con la máxima seriedad y de la forma que creyesen más conveniente (aunque no debían ser desproporcionadas).

A los participantes seleccionados para desempeñar el papel de prisioneros se les dijo, simplemente, que esperasen en sus casas a que se les “visitase” el día que empezase el experimento. Sin previo aviso fueron “imputados” por robo a mano armada y arrestados por policías reales del departamento de Palo Alto, que cooperaron en esta parte del experimento. Pasaron un procedimiento completo de detención policial, incluyendo la toma de huellas dactilares y una fotografía para ser fichados, se les leyeron sus derechos... Después fueron trasladados a la prisión ficticia (con los ojos vendados), donde fueron inspeccionados desnudos, “despiojados”, y se les dieron sus nuevas identidades.

Tras un primer día de relativa tranquilidad, el segundo día se desató el primer motín de los presos (ante las primeras muestras de control de los guardias). La situación provocó una fuerte discusión entre los guardias sobre el grado de dureza a aplicar. Los guardias hicieron horas extras para disolver la revuelta, atacando a los prisioneros incluso con extintores sin la supervisión directa del equipo investigador. A partir de entonces trataron de dividirlos en “buenos” y “malos”, y de enfrentarlos (haciéndoles creer que había “informantes” entre ellos): a unos presos se les imponían tareas humillantes y a otros se les concedían privilegios (a quienes demostrasen buen comportamiento). Los “recuentos” de prisioneros, que habían sido ideados inicialmente para ayudar a los prisioneros a familiarizarse con sus números identificativos, evolucionaron hacia experiencias traumáticas en las que los guardias atormentaban a los prisioneros y les imponían castigos físicos, que incluían ejercicios forzados. El derecho de ir al lavabo pasó a ser un privilegio, que podía ser denegado. Se obligó a algunos prisioneros a limpiar retretes con sus manos desnudas. Se retiraron los colchones de las celdas de los “malos” y también se forzó a los prisioneros a dormir desnudos en el suelo de hormigón. La comida se negaba con frecuencia

como medida de castigo. Además, se les obligó a ir desnudos y a llevar a cabo actos homosexuales como humillación.

Uno de los presos simulados, el 8612, comenzó a mostrar síntomas de ansiedad, trastorno emocional agudo, razonamiento ilógico, llanto incontrolable y ataques de ira, expresando al consultor principal del experimento su voluntad de abandonarlo. Pero, pensando que era una excusa para dejarlo injustificadamente, no le autorizaron a hacerlo (en contra de lo que se les había informado). Esta impresión se extendió entre los restantes presos, que dejaron de percibirlo como un experimento, o un modo de conseguir dinero, y empezaron a contemplarlo como una experiencia “real” en la que ellos eran “auténticos” presos, del mismo modo que ocurría con los carceleros. El estado psicológico del recluso 8612 empeoró y parecía estar fuera de control, lo que llevó a los investigadores a ponerlo en libertad. Tras su liberación, otros reclusos simulaban síntomas psicológicos parecidos para que les liberasen, mientras que otros optaron por mostrar un comportamiento lo más correcto posible para evitar males mayores.

En el cuarto día, Zimbardo y los guardias reaccionaron ante el rumor de un plan de huida intentando trasladar el experimento a un bloque de celdas reales en el departamento local de policía porque era más “seguro”. Pero la Policía rechazó su petición. Aunque la huida finalmente no tuvo lugar, el rumor sirvió para que los carceleros guardias desconfiasen de los presos, a quienes fueron tratando con creciente dureza física y psicológica. Ello propició la reacción violenta de los reclusos, y la aún más violenta contrarreacción de los guardias, derivando en un clima de escalada máxima que obligó a la dirección científica de la *simulación* a detener el experimento, cuando únicamente se había alcanzado el sexto día del mismo. En menos de una semana, 24 jóvenes universitarios, absolutamente sanos y normales, se convirtieron en guardias todopoderosos y sádicos (particularmente por la noche, cuando pensaban que las cámaras estaban apagadas), y en presos desvalidos e incluso algunos con alteraciones psíquicas de diversa consideración. Muchos de los guardias se enfadaron cuando el experimento fue cancelado.

Este experimento se utilizó para ilustrar la teoría de la *disonancia cognitiva* y el *poder* de la autoridad: la obediencia cuando se proporciona una ideología legitimadora y el apoyo institucional. También tuvo implicaciones dogmáticas en el ámbito del derecho penal (Gómez, 2006). Pero fue ampliamente criticado por:

- Su falta de ética y el hallarse en los límites del *método científico*. Fromm (1973) cuestionó que sus resultados pudiesen generalizarse y que no se hubiesen realizado los “controles científicos” tradicionales. Para Oskamp (1977), es un estudio de *simulación* (porque se imitan algunos aspectos cruciales de una situación del mundo real) y se hace con el conocimiento de los sujetos que participan en la *simulación*.
- Los participantes basaban su conducta en cómo se esperaba que se comportasen y la modelaron de acuerdo con estereotipos que ya tenían sobre prisioneros y guardias (como si fuera un *juego de rol*). Para Zimbardo los participantes internalizaron sus papeles a medida que el experimento continuó.
- Falta de *validez ecológica*. Muchas de las condiciones que se impusieron fueron arbitrarias y distantes de la “realidad” de las prisiones (como, *por ejemplo*, la llegada de los prisioneros con los ojos vendados, que vistieran sólo con batas, sin ropa interior, que no pudieran mirar a través de ventanas, o prohibirles usar sus nombres reales).
- *Determinismo*: los informes describen diferencias significativas en la crueldad de los guardias (algunos fueron más amables), pero no se explicaron.

- Se observó una *muestra* muy pequeña, de sólo 24 participantes, en un corto período de tiempo, y todos interactuaban en *un mismo grupo*.
- Se concedió una gran libertad de acción a los sujetos experimentales. La interpretación de los resultados habría sido más inequívoca con un mayor *control experimental*.

Inevitable fue su comparación con el experimento también famoso realizado en 1961 en la Universidad de Yale por Stanley Milgram, un antiguo amigo de Zimbardo, y publicado en 1963 (“Behavioral Study of Obedience”. *Journal of Abnormal and Social Psychology*). Este último experimento analiza la disposición a obedecer las órdenes de una autoridad, aun cuando éstas pudieran entrar en conflicto con la conciencia personal. El investigador persuadía al participante para que diera lo que éste creía eran descargas eléctricas dolorosas a otro sujeto (que simula recibirlas). Pese a sus súplicas para que no lo hiciesen, muchos participantes continuaron dando descargas eléctricas.

3.4. Amenazas a la validez del experimento

El ejemplo anterior da pie a señalar las principales amenazas a la “validez” de los experimentos. El análisis *causa-efecto*, que caracteriza a la investigación experimental, exige un riguroso *control* del proceso de investigación. *Control* que se dirige a la eliminación de explicaciones alternativas que invaliden las conclusiones del estudio. A lo ya señalado en el capítulo 1, referido a las distintas *fuentes de invalidez* a partir de las aportaciones de Campbell y Stanley (1963, 1970) y Cook y Campbell (1979), se añaden otras varias también a ser controladas en busca de *validez interna*. Las *fuentes de invalidez* principales son:

- a) *Maduración*: las personas cambian con el paso del tiempo, en sus formas de pensar y de actuar. Este cambio (o *maduración* biológica, psicológica, emocional...) puede incidir en los resultados del *experimento*. De manera especial, cuando se analiza a unos mismos sujetos durante un largo período de tiempo (diseños *intrasujetos*). Los cambios registrados entre la medición previa y posterior al *tratamiento* pueden erróneamente atribuirse a su efecto, cuando en realidad se deben a la propia evolución del sujeto que se observa.
Esta fuente de invalidez puede controlarse mediante un *grupo de control*. Partiendo de una *maduración* similar de los sujetos en ambos grupos (*experimental* y de *control*), las diferencias en los valores medidos de la variable dependiente podrán atribuirse al efecto del *tratamiento*, y no a la evolución de los sujetos experimentales.
- b) *Historia*: acontecimientos externos (ajenos al experimento) que pueden afectar a sus resultados. Es más probable que acontezca en experimentos de larga duración.

La generación de un *grupo de control* también ayuda a “controlar” su efecto, dado que su presencia será igual en ambos grupos.

- c) *Instrumentación*: las variaciones en las mediciones de la variable *dependiente* antes y después del *tratamiento* experimental pueden deberse a cambios en el instrumento de medición (escala, cuestionario...). No puede haber variaciones importantes ni en el contenido ni en la forma de administrarse.

La existencia de un *grupo de control* permite asimismo contrarrestar los efectos debidos al instrumento de medida. Al ser el mismo, en los diferentes grupos constituidos, las diferencias en sus resultados podrán atribuirse al *tratamiento* dado.

- d) *Mortalidad experimental*: durante el curso del experimento, algunos sujetos de la *muestra* pueden dejar de participar. Su abandono puede repercutir en los resultados de la investigación. Más cuanto mayor disparidad exista entre las características de quienes abandonan el experimento y quienes aún permanecen en él. Alterará la composición inicial de la *muestra*, con su consiguiente pérdida de representatividad.

Esta fuente de invalidez también es más probable conforme aumenta la duración del experimento. El tiempo de duración propicia la aparición de factores *internos* (cansancio, pérdida de interés) y *externos* (cambio de residencia, ocupación, o cualquier cambio en la forma de vida del sujeto que dificulte su continuidad en el experimento), que provocan la *mortalidad experimental*.

El uso de *varios grupos de tratamiento*, y la *comparación* de sus resultados, puede también ser de utilidad para controlar esta cuarta fuente de *invalidez*.

- e) *Efecto de la regresión*: algunos sujetos, que partieron de posiciones muy extremas en la variable *dependiente*, pueden modificar su posición en la segunda medición y no como consecuencia del *tratamiento*, sino de la tendencia observada a dar puntuaciones menos extremas en mediciones posteriores.

Si se asignan los sujetos con posiciones iniciales extremas al azar a los diferentes grupos, las oscilaciones en sus puntuaciones serán iguales en todos los grupos (con independencia del *tratamiento*). Ello posibilita la atribución de la disimilaridad de los resultados al efecto del *tratamiento*.

- f) *Efecto de la medición inicial*: el *pretest* (o medición previa de la variable *dependiente*) puede orientar a los sujetos experimentales sobre la finalidad del estudio y lo que de ellos se espera. Esto puede afectar a que modifiquen su actuación, influyendo en los resultados del estudio. El *pretest* también favorece la familiarización con el *instrumento* de medición. Lo que puede contribuir a que el individuo mejore su puntuación (en una segunda medición), independientemente del *tratamiento* dado.

La creación de un *grupo de control*, igualmente sometido a mediciones previas y posteriores al *tratamiento*, también contribuye a descartar esta fuente de *invalidez*. Como en todos los grupos ha habido una medición de la

variable *dependiente* anterior al *tratamiento*, las diferencias que se observen tras la manipulación experimental se podrán atribuir al *tratamiento* dado.

De lo expuesto se colige que estas seis fuentes principales de *invalidéz* pueden “controlarse” mediante la creación de un *grupo de control*. Si bien la existencia de este grupo puede a su vez provocar otras *fuentes de invalidéz* por la existencia de:

- *Sesgos en la selección y asignación* de los sujetos a los grupos *experimental* y de *control*, que contribuyan a que los perfiles de los integrantes de cada grupo difieran antes de aplicar el *tratamiento* experimental. Lo que invalidaría los resultados atribuidos al efecto diferencial del *tratamiento*. La asignación aleatoria de los sujetos a los grupos contribuye a mermar este *sesgo*.
- *Compensación al grupo de control por la privación del tratamiento* (si éste se estima de valor). Dicha *compensación* puede ser un trato más favorable por parte del personal de la investigación, una mejor ubicación física del grupo, o cualquier otra que reste eficacia al *grupo de control* (para eliminar explicaciones alternativas al *tratamiento*). En este caso, como en el anterior, el *grupo de control* que resulta no es “auténtico” y no actúa como tal.

Por otra parte, el conocimiento, por parte de los integrantes del grupo de *control*, de la privación de *tratamiento*, puede afectar su intervención posterior en la investigación porque:

- Se *desmoralicen* y suspendan su actuación en el experimento o bajen su actuación.
- *Rivalicen con el grupo experimental*, intentando su superación.
- *Imiten el tratamiento* dado al grupo *experimental* (manifestando sus mismas actitudes y conductas).

El conocimiento y reducción de estas fuentes diferentes de *invalidéz interna* redundan en la reducción de *explicaciones alternativas* a la *manipulación experimental* y la consiguiente ganancia en *validez interna*.

Respecto a la *validez externa* (o capacidad de generalización de los resultados del experimento) ha de señalarse:

- a) El bajo tamaño muestral y su autoselección actúan en su contra. *Por ejemplo*, la práctica habitual de escoger alumnos universitarios como sujetos experimentales invalida las posibilidades de generalizar los hallazgos al conjunto de la población.
- b) La falta de *realismo* también limita la posibilidad de generalización a la vida diaria, a partir de un *tratamiento* “artificial”.

Por ejemplo, el efecto de ver un videojuego violento durante una hora en un aula no es generalizable al efecto de su visión durante muchas horas y días de un

año. En la medida en que el experimento alcance “realismo”, sus hallazgos alcanzarán *validez externa*.

- c) Existencia de “*reactividad*”: que los sujetos no actúen como lo harían en la vida real porque saben que están participando en un experimento. Ésta puede deberse a la confluencia de distintos efectos (algunos de ellos ya comentados).
 - *Efectos de Hawthorne*, tipificado por Roethlisberger y Dickenson (1939) a partir del experimento de *Hawthorne*. Indistintamente de la condición laboral modificada, la productividad se elevaba porque los trabajadores no respondían al *tratamiento*, sino a la atención adicional que recibían por ser parte de un experimento y saber que estaban siendo observados.
 - *Efecto de las características de la demanda* de Orne (1962): los sujetos experimentales ajustan su conducta a lo que se piensa que se demanda de ellos (apoyo de las hipótesis de la investigación), con objetivos de agradar al investigador.
 - *Efecto placebo*: a los sujetos se les da el “placebo” y actúan como si hubiesen recibido el tratamiento verdadero (Neuman, 2000).

Por ejemplo, administrarles agua en vez del fármaco que ellos esperan y actuar como si hubiesen recibido el susodicho medicamento o *tratamiento*.

- *Efecto novedad*: el efecto de algo nuevo puede desaparecer a lo largo del tiempo (Smith y Glass, 1987).

Es bueno que al final del experimento haya un *feed-back* entre el investigador y los sujetos experimentales. Dar oportunidad para que los primeros expliquen los propósitos reales de la investigación y respondan a las preguntas que los segundos les planteen. Por su parte, los participantes en el experimento podrán explicar su experiencia, cómo lo vivieron, qué cambios observaron en su conducta y a qué lo atribuyen. Sin duda esto ayudará en la interpretación de los resultados del experimento.

3.5. Ventajas e inconvenientes de la experimentación como estrategia de investigación

El cuadro 3.3 sintetiza las *ventajas e inconvenientes* principales de la *experimentación* como estrategia de investigación social. De ellas destaca, en el lado positivo, su adecuación para el estudio y verificación de relaciones causales, además del análisis de los procesos de interacción en grupos pequeños. A ello se suma su reducido coste económico (en relación con otras estrategias de investigación cuantitativa, como la *encuesta*), porque exige menos personal y tamaño *muestral*. Además está la facilidad de su *replicación*, destacada por Babbie (1992).

En su contra actúa su referida “artificialidad”, en especial de los experimentos de *laboratorio*. También que los sujetos suelen ser conscientes de que se les está observando y ello puede afectar a su proceder en el experimento. Este inconveniente, junto a los debidos a la *muestra* analizada (tamaño y procedimiento de selección de los sujetos en la población, rara vez a través de métodos aleatorios), limita las posibilidades de inferencia y generalización de los resultados de la investigación.

“Muchos adultos simplemente no están disponibles para participar en experimentos, y los voluntarios no son representativos de la generalidad de la población. Esto limita la investigación experimental a poblaciones cautivas (como estudiantes, delincuentes recluidos en instituciones correccionales, trabajadores de fábricas, miembros de las fuerzas armadas)” (Hakim, 1994: 102).

La investigación experimental suele circunscribirse a temas específicos y a variables que puedan manipularse. Cuestiones éticas impiden la *manipulación experimental* de muchas variables de interés para el investigador social.

Por ejemplo, no sería ético separar a unos niños pequeños de su entorno para, con posterioridad, comparar su desarrollo socio-afectivo con el de aquellos niños que permanecieron en su propio entorno sociofamiliar. Tampoco sería admisible, desde el punto de vista ético, la asignación aleatoria de niños huérfanos a hogares caracterizados por experiencias de violencia en distintos niveles (con la finalidad de medir la influencia de los malos tratos en la posterior evolución conductual de los niños).

“El objetivo fundamental de un experimento es conseguir el máximo impacto en los participantes dentro de los límites éticos y la necesidad de control de sus variables” (Aronson *et al.*, 1985: 482).

CUADRO 3.3. Ventajas e inconvenientes de experimentación como estrategia de investigación

VENTAJAS	INCONVENIENTES
<ul style="list-style-type: none"> – Permite verificar relaciones causales – Favorece el control a priori de explicaciones alternativas a las relaciones <i>causa-efecto</i> analizadas (<i>validez interna</i>) – Ayuda a los análisis de los procesos de interacción grupal – Facilidad de repetición de la investigación – Reducido coste económico (comparado con otras estrategias de investigación) 	<ul style="list-style-type: none"> – Límites a la inferencia y generalización de los resultados (<i>validez externa</i>). Las exigencias de “control” propicia su aplicación en entornos artificiales (laboratorios), a grupos pequeños, y a unas cuantas variables – Utilizar <i>muestras cautivas</i> (personal voluntario) restringe la representatividad de la muestra – Difícil materialización en entornos naturales – Efectos reactivos por la presencia del investigador

Lecturas complementarias

- Alvira, F. *et al.* (1981): *Los dos métodos de las ciencias sociales*. CIS Madrid.
- Arnau, J. (1984): *Diseños experimentales en psicología y educación*. Trillas. México.
- Campbell, D. y Stanley, J. (1970): *Diseños experimentales y cuasiexperimentales en la investigación social*. Amorrortu. Buenos Aires.
- Campbell, D. T. y Russo, M. J. (1999): *Social experimentation*. Sage. Thousand Oak, CA.
- Cook, T. y Campbell, D. (1979): *Quasi-experimentation: design and analysis issues for fields settings*. Rand McNally. Chicago.
- Morales Domínguez, J. *et al.* (1981): *Metodología y teoría de la psicología*. UNED. Madrid.
- Neuman, W. L. (2000): *Social research methods: qualitative and quantitative approaches*. Allyn and Bacon. Boston, MA.
- Orr, L. L. (1999): *Social experiments: evaluating public programs with experimental methods*. Sage. Thousand Oaks, CA.
- Saxe, L. y Fine, M. (1981): *Social experiments. Methods for design and evaluation*. Sage. Beverly Hills.

Ejercicios propuestos

1. De cada uno de los experimentos ilustrados a lo largo del capítulo, describa qué tipo de experimento es y si cumple todos los criterios de validez. En caso negativo, ¿qué debería haberse hecho?
2. Ilustre las diferentes fuentes de invalidez de un experimento con un ejemplo.
3. Describa un experimento publicado en una revista científica, desde sus inicios hasta su conclusión. Especifique si cumple los diferentes criterios de validez.
4. ¿Qué experimento diseñaría para comprobar los efectos del consumo de alcohol en la desinhibición. Destaque posibles variables “perturbadoras” y actuaciones para su control.

4

LA INVESTIGACIÓN “NO REACTIVA” A PARTIR DE DATOS DISPONIBLES

Los *experimentos*, al igual que la *encuesta*, son estrategias de investigación “reactivas”. La presencia del investigador puede afectar a la calidad de las respuestas, como muestran los capítulos 3 y 5: efectos debidos a su actuación e, incluso, a sus características personales. Además afecta que la persona se sienta observada y ello le lleve a alterar su conducta, actitud u opinión, para adecuarla a lo que piensa se espera de ella. Recuérdense los efectos de *Hawthorne* y de las *características de la demanda* de Orne (1962), ya referidos en el apartado 3.4.

En los capítulos precedentes también se ha resaltado la relevancia de la *revisión bibliográfica* para la configuración del *marco teórico* de la investigación. La obtención de información a través de fuentes bibliográficas se encuadra dentro de lo que se entiende por “*investigación no-reactiva*”. Incluye la recogida de información a partir de los datos que ofrecen *archivos documentales y estadísticos*, *documentos personales* (autobiografías, diarios, memorias, cartas, fotos o cualquier registro iconográfico u objeto personal) y la *observación no interactiva*. Indaguemos en qué consiste, en su aplicación en la investigación cuantitativa. Para su empleo en la investigación cualitativa, se remite a manuales especializados, como el publicado en esta misma editorial por Valles (1997/2009).

4.1. La medición “no reactiva”

El término *unobtrusive measures* fue acuñado en la monografía que Eugene Webb, Donald Campbell, Richard Schwartz y Lee Sechrest publican en 1966, con el título: *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. En ella introducen al lector al análisis de estudio de huellas o rastros físicos, de archivos estadísticos y documentales, y a la observación, como medios no intrusivos para analizar la

realidad social. Se contraponen a las mediciones obtenidas mediante *entrevista*, *encuesta* o la *experimentación*, donde los sujetos son conscientes de que son objeto de estudio, lo que les hace comportarse de manera “no natural”. Si bien pueden utilizarse de manera complementaria. Los autores destacan la importancia de “triangular” los resultados obtenidos a través de estrategias o métodos diferentes para contrapesar los límites y sesgos a los que cada uno es vulnerable.

El análisis se circunscribe a lo observado en entornos naturales (ajenos a la manipulación experimental), a lo que queda registrado sin conocimiento de que se es objeto de estudio. La información se infiere a partir de los *datos disponibles*. Se diferencia un amplio abanico de medidas “no reactivas”:

- a) *Rastros físicos*: evidencias “físicas” a través de las que se infiere actividad o conducta social.

Por ejemplo, el desgaste de las baldosas como evidencia de los sitios más populares o de mayor concurrencia de personas; las emisoras que se graben en la radio de los coches como muestra de la personalidad de su usuario; o el análisis de los grafitis de los aseos públicos. También se puede extraer mucha información de los datos registrados en las lápidas (fecha de nacimiento, defunción, matrimonio, descendencia...). Muestra el estudio de Foster, Hummel y Adamchak (1998), destacado por Neuman (2000): extrajeron datos sobre las fechas de nacimiento y defunción de 2.000 tumbas de 10 cementerios del área de Illinois, desde el período de 1830 a 1989. Observaron que las mujeres de entre 10 a 64 años tenían una mayor probabilidad de fallecimiento que los varones, y que la población más joven moría a finales de verano, mientras que la más mayor a finales de invierno.

- b) *Archivos* (bibliográficos, documentales, estadísticos, de prensa...) de fuentes públicas y privadas, y de materiales escritos y audiovisuales. Limitado es el análisis de la realidad social sin el recurso a fuentes documentales y estadísticas.

Muestra el estudio clásico sobre el *Suicidio* de Emile Durkheim, de 1897, a partir de las *estadísticas* disponibles; o el también clásico de William I. Thomas y Florian Znaniecki *El campesino polaco en Europa y América*, de 1818-1920, donde la integración de los inmigrantes polacos en EE UU se analiza conjugando información proporcionada de diferentes fuentes: 754 cartas de inmigrantes polacos, archivos de un periódico adquirido por Thomas en Polonia (y que cubría un período de veinte años), archivos de parroquias polacas en EE UU, documentos de organizaciones de ayuda a inmigrantes polacos, más la autobiografía de un joven inmigrante polaco llamado Wladeck. En la actualidad, la proliferación de medios digitales y el uso de Internet facilita el acceso y el análisis de datos de *archivos* (nacionales e internacionales).

- c) *Observación*: puede ser con participación o no del investigador, e incluso en el marco de un experimento. La condición es que su presencia esté camuflada y que los sujetos no se sientan observados. Esta “ignorancia” lleva a que

se comporten de forma “natural”, sin sesgar los resultados del estudio, aunque limita el uso de cámaras fotográficas, de vídeo, grabadoras... En caso de ocultación, están los límites legales a su uso, sin el consentimiento previo de los sujetos observados. Cualquier conducta externa puede ser objeto de observación y cuantificación.

Por ejemplo, observar si los conductores se paran ante una señal de stop, durante un periodo de tiempo (durante el día, la noche); anotar las características del vehículo, del tráfico y los conductores, que expliquen el grado de cumplimiento de la obligación de parar el vehículo al llegar a la señal de stop.

La ventaja principal de la investigación *no intrusiva* (o reactiva) es que mide la conducta “real” como opuesta a la “autoinformada” (a lo que se dice que se piensa o hace a través de entrevista, cuestionario, grupo de discusión...). Y sobre todo su bajo coste económico-temporal. En su contra, que la observación de la conducta “real” no necesariamente es unívoca. Diferentes investigadores pueden fijarse y destacar aspectos distintos de lo que observan. De ahí la conveniencia tanto de la “triangulación metodológica” como de “investigadores”, como se vio en el capítulo 1.

4.2. Investigar a partir de fuentes secundarias: archivos bibliográficos, documentales, estadísticos, de prensa

Rara vez la investigación queda circunscrita al análisis de los datos generados por el propio investigador (investigación *primaria*). Su información precisará compararse con la obtenida por otros investigadores (*revisión bibliográfica*) y otras fuentes de información *secundaria* (de tipo estadístico, documental, de prensa...). “*Secundaria*” porque no ha sido producida por el mismo investigador y con una finalidad que no tiene por qué ser coincidente con los objetivos de su investigación.

Los hallazgos de un experimento o encuesta propia, *por ejemplo*, pueden contrastarse con los obtenidos por otros investigadores, y que están disponibles en archivos o bancos de datos. También pueden contextualizarse en una coyuntura socioeconómica concreta a partir de estadísticas elaboradas por organismos oficiales. Sin duda la conjugación de diferentes fuentes de información enriquece la investigación, como se vio en el capítulo 1.

Pero, al igual que en la investigación *primaria*, la *secundaria* puede quedar restringida al análisis de datos disponibles y no por ello carecer de relevancia. El término “*secundaria*” no implica nada sobre la importancia de la investigación; sólo que se asienta en datos no generados por el propio investigador. Como señala Stewart (1984: 13), “quizás es desafortunado que el término *secundaria* se haya elegido para referirse a datos existentes”.

4.2.1. Revisión de investigaciones teórico-empíricas publicadas

En el capítulo 1 se destacó la relevancia de la *revisión bibliográfica*, de investigaciones teóricas y empíricas, como paso previo y fundamental en la configuración del *marco teórico* de cualquier investigación. Contribuye a:

- a) *Familiarizarse* con el tema de estudio: antecedentes, aspectos analizados, metodología aplicada, resultados obtenidos.
- b) *Elaborar* una *síntesis* del conocimiento existente que permita el estudio dentro de una trayectoria concreta de investigación. Síntesis que resulta de la valoración de la información disponible y su adecuación a los propósitos de la investigación.
- c) *Estructurar* las ideas originarias del estudio en un *diseño de investigación* específico. De la *revisión* de estudios teóricos y empíricos sobre el tema de interés puede extraerse información que contribuya a concretar las *hipótesis* a comprobar, los sujetos a analizar, junto a las *estrategias* y *técnicas* de recogida y de análisis a aplicar (a partir de los resultados y experiencias de estudios previos).

La *revisión de investigaciones* varía en *énfasis* (teórico o metodológico) y en el período de *tiempo* que comprenda. Incluye sólo estudios coetáneos o puede abarcar diferentes momentos temporales, proporcionando una perspectiva histórica a la investigación. El número de estudios revisados varía en función del *tema* y la *experiencia* del investigador en la localización del material relevante (Hakim, 1994). También de su interés en profundizar en el conocimiento ya existente.

Sin duda Internet ha ampliado las posibilidades de acceso a la información disponible en bibliotecas, editoriales, revistas especializadas... de los diferentes países del mundo; además de poder acceder a los textos en formato digital. Tradicionalmente la manera más rápida de acceder a investigaciones publicadas había sido a través de *fuentes de referencias* como:

- *Social Sciences Citation Index* (SSCI) –que compila unos 130.000 artículos de revista cada año–, y otros índices que cubren bibliografía especializada: *Current Index to Journals in Education*, *Index of Economic Articles*, *Population Index*, *International Population Census Bibliography*, *Scopus*, *Latindex*, *JSTOR*, *Socindex and Socindex with Full Text*, *Academic Search Complete*, *ISOC* (Índice Español de Ciencias Sociales y Humanidades), *PIO* (Periodical Index Online) o *Dialnet*, entre otros.
- *Abstracts* (Sociological Abstracts, Psychological Abstracts, Abstracts in Anthropology, International Dissertation Abstracts, International Political Science Abstracts, Linguistic and Language Behaviour Abstracts, Social Service Abstracts...): contienen breves resúmenes (en torno a 100 palabras redactadas por el propio autor del estudio) de investigaciones publicadas en diferentes revistas internacionales. Las investigaciones figuraban ordenadas por tema y autor.

- *Reviews* (Sociological Review, Review of Educational Research, Psychological Bulletin...): comprenden artículos y reseñas de publicaciones (libros y revistas) sobre un tema particular.

Reseñas de libros también pueden encontrarse en la generalidad de las revistas científicas editadas, aunque son pocos los libros que acaban siendo referenciados. Asimismo se recomienda no limitarse a los artículos que aparezcan reseñados en los *indexes* o *Abstracts*. Además de no abarcar todo lo editado en revistas de impacto (difícil de cubrir en tiempos, como los actuales, de amplia publicación de estudios), se enfrentan a tres *límites* importantes, ya destacados por Cooper (1984):

- La *carencia temporal* (con frecuencia superior a los dos años) desde que el estudio concluye y queda registrado en el *abstract*.
- La *división en áreas concretas*, cuando las investigaciones con frecuencia son interdisciplinarias.
- Su *organización*. Aun recogiendo lo publicado en revistas de impacto, es probable que no llegue a accederse a algunos artículos porque no sean localizados con facilidad. Esto último ocurre cuando el investigador no conoce términos claves de los *índices* que se aplican a artículos, o cuando el autor de los *índices* omite términos que el investigador emplea comúnmente.

La última limitación es extensible a cualquier buscador o localizador de información, como *Google*, *Yahoo* o *Ask.com*, por ejemplo. Aunque ayuden a la búsqueda inicial de información, el acceso directo a las páginas webs de las revistas especializadas (como la *REIS* –Revista Española de Investigaciones Sociológicas–, *Social Science Research*, *Public Opinion Quarterly*, *European Sociological Review*, *Quality & Quantity*...) es imprescindible para alcanzar una mayor cobertura de lo publicado. En la actualidad no se precisa su búsqueda a través de bibliotecas, a lo archivado en formato papel o digital. Puede accederse directamente a sus páginas webs, donde no sólo se detalla el índice o contenido de los últimos números publicados (y de los anteriores en su archivo digital), y al resumen de los artículos, sino que también facilitan el acceso (incluso gratuito) al contenido íntegro del artículo, proporcionando al investigador inmediatez en el acceso a la información.

En España, el *Instituto de Información y Documentación en Ciencias Sociales y Humanidades* (ISOC), dependiente del CSIC (el Consejo Superior de Investigaciones Científicas) dispone de un servicio especializado en información bibliográfica. Este servicio comprende la edición de repertorios bibliográficos y la creación de *bases de datos* automatizadas.

A través de los artículos y libros localizados puede asimismo accederse a otras investigaciones referenciadas en sus bibliografías. Éstas remiten a su vez a otras, y así como si de una *bola de nieve* se tratase, hasta que el investigador decida dar por concluida la *revisión bibliográfica*. Ello puede deberse a que se ha alcanzado el principio de *saturación*: las últimas publicaciones a las que accede no proporcionan

ningún dato nuevo, sino que redundan en la información que ya tiene registrada. O, simplemente, porque piensa que ha cubierto todos los aspectos de interés o no desea destinar más tiempo y recursos a la búsqueda de información bibliográfica.

Por último, los congresos, seminarios o simposios, permiten el conocimiento y divulgación de investigaciones *no publicadas* y realizadas en fechas recientes. Estas investigaciones figuran reseñadas en los *resúmenes* de comunicaciones y ponencias que edita el organismo responsable del congreso o simposio. Véase, por ejemplo, los publicados por la *FES* (Federación Española de Sociología), la *ESA* (European Sociological Association) o la *ISA* (International Sociological Association) de los últimos congresos celebrados.

4.2.2. Análisis de datos secundarios

La investigación *secundaria* no queda restringida a la revisión, la síntesis y el análisis de investigaciones publicadas. Incluye el análisis de investigaciones no publicadas, de estadísticas e informes elaborados por agencias oficiales o institutos estadísticos, sin olvidarnos de los datos “brutos” que cada organismo o empresa recoge relativos a su personal y actuación.

- a) Toda empresa u organismo registra información sobre su personal y actividad. Información que resulta imprescindible para describir y evaluar su actuación.

Por ejemplo, los *expedientes* incoados por instancias judiciales, los *informes* de pacientes de centros sanitarios, los *expedientes académicos* de alumnos de centros de enseñanza, los *registros* de afiliados a asociaciones o del personal de una empresa.

El carácter privado (no “público”) de este tipo de información obstaculiza su acceso. En caso de precisarla, el investigador tendrá que solicitarla, expresa y formalmente, al organismo en cuestión. La solicitud ha de enfatizar la trascendencia de la información requerida para cubrir los objetivos de la investigación; además de garantizar el anonimato de las personas a las que se refieran los datos manejados.

EJEMPLO DE INVESTIGACIÓN SECUNDARIA A PARTIR DE EXPEDIENTES

Un ejemplo lo proporciona la tesis doctoral de M.^a Á. Cea D’Ancona, *La Justicia de Menores en España. Funcionamiento y resultados*, leída en junio de 1990 y publicada, primero, en facsímil por la Editorial de la Universidad Complutense de Madrid en 1991, y por

el CIS/Siglo XXI en 1992, en la colección *Monografías*, n.º 127 (con el título *La justicia de menores en España*).

La investigación tenía por *objetivo* principal evaluar la actuación de los (ya extinguidos) tribunales tutelares de menores. Quería comprobarse si el sistema de justicia de menores cumplía sus propósitos originarios (rehabilitar al transgresor, reeducándolo y adaptándolo a la vida social) o, por el contrario, producía *resultados* no deseados contribuyendo al desarrollo de la “carrera delictiva”, como afirman los teóricos del *Etiquetamiento*, como Howard Becker o Edwin Lemert.

Primero se realizó una *evaluación del proceso* (o *formativa*), con la finalidad de describir el sistema de justicia de menores en España (en la teoría y en la práctica), a nivel normativo y organizacional. Una vez conocido su funcionamiento real (primer objetivo), se comprobó su efectividad o *impacto* (segundo y principal objetivo de la investigación). Ambos objetivos se ajustan a dos modalidades de *evaluación*: del *proceso* (o *formativa*), y de *impacto* (o *sumativa*), descritas en el capítulo 1. A partir de ambas evaluaciones, se deducen los aspectos más necesitados de reforma en la justicia de menores (que era el tercer objetivo específico de la investigación).

Para el cumplimiento de los objetivos marcados, se diseñó una investigación plenamente “*secundaria*”, que comprendía tres fases principales (cuyos pasos detalla el *Apéndice Metodológico* incluido al final de la monografía):

1. *Búsqueda de investigaciones* teórico-empíricas publicadas sobre justicia y delincuencia juvenil, *evaluación de programas*, técnicas de recogida y análisis de información (para poder abordar la parte empírica de la investigación). A partir de la *revisión bibliográfica*, se concretan los *objetivos* y el *marco teórico* de la investigación.
2. La *evaluación del proceso* incluye:
 - *Documentarse* sobre el fundamento y el origen de la justicia de menores, instituciones y agentes colaboradores. La lectura comienza por el texto refundido de la legislación de los tribunales tutelares de menores vigente en España en el momento de la investigación (la *LTTM* de 1948). Prosigue con varios estudios de derecho comparado de menores en España, Alemania Federal, Francia, Inglaterra, Italia y Estados Unidos. De estos estudios se extrajeron los puntos que diferencian y asemejan a los diversos sistemas legislativos en materia de creación, organización, competencia y funcionamiento de los tribunales e instituciones auxiliares. También, su concepción de la conducta delictiva y de la minoría de edad penal, como fundamento y razón de ser de la jurisdicción de menores.
 - *Analizar la práctica real* de los tribunales tutelares de menores a partir de la información que ellos registran. Principalmente los *expedientes* de menores incoados en el Tribunal Tutelar de Menores de Madrid y la información estadística recopilada por los tribunales.

Para la selección de los *expedientes* se conjugan tres criterios que determinaron la *muestra* de la investigación:

- Como el objetivo de la investigación era evaluar la intervención del tribunal, se descartó la población cuya actividad delictiva “registrada” fue episódica

- (al menos hasta que alcanza la mayoría de edad penal). Interesaban los menores que, bien por sus características personales o por la entidad de los delitos cometidos (o denunciados), quedaban bajo la tutela del Tribunal.
- Como *evaluar el impacto* de un *programa* precisa que la intervención del *programa* haya concluido, y no en fecha reciente (al menos hace dos años), se decidió excluir del estudio *expedientes* de menores abiertos o cerrados con posterioridad a 1986 (fecha de comienzo de la investigación, que duró cuatro años).
- La *muestra* debía ser amplia y lo más representativa posible de la población de estudio, para que la investigación alcanzara *validez externa*. Además debía abarcar diferentes momentos históricos vividos en fechas próximas y que, probablemente, habrían repercutido en la política de tratamiento y prevención de la delincuencia juvenil.

Con estos requisitos en mente, se consultaron las *estadísticas propias* del Tribunal de Menores de Madrid, referentes a su actuación desde 1975 hasta 1983 (de fechas posteriores se carecía de información estadística). En las estadísticas se observaron cambios de tendencia en la política de tratamiento de la delincuencia desde 1975. Por lo que se incluyeron todos los casos con *expedientes* abiertos en 1975, 1977, 1979, 1981 y 1983. Estos debían también reunir la condición de haber estado bajo tutela del tribunal con anterioridad a 1986.

La *muestra* final la forman 715 *expedientes* de menores que habían estado en libertad vigilada o en internamiento antes de 1986; 90 *expedientes* quedaron excluidos bien porque no se encontraban en el *Archivo* de expedientes durante el tiempo que duró la recogida de información (porque aún estaban abiertos), bien porque sus datos no se correspondían con los que figuraban en el *Libro de Actas*. Y a la inversa, había otros *expedientes* que sí cumplían los criterios marcados para formar parte de la *muestra* y, en cambio, no figuraban en el *Libro de Actas*. Errores de registro que determinó que, en vez de ir a un *expediente* específico, se revisaran uno a uno todos los abiertos en las fechas indicadas y que, al estar ya cerrados, se encontraban en el *Archivo*.

En el despacho del psiquiatra (que sólo acudía al tribunal una o dos veces a la semana), la autora estuvo durante un año leyendo cada uno de los 715 *expedientes* (que iba cogiendo del *Archivo*). Cuando el psiquiatra se personaba, el pequeño habitáculo donde se guardaban los últimos *expedientes* se convertía en sala de lectura.

De los primeros *expedientes*, aleatoriamente escogidos, se extrajo toda la información que se estimó de relevancia para cubrir los objetivos de la investigación. Al leerlos descubrió la riqueza de información que contenían los diversos *informes adjuntos al expediente* (de la Policía, del servicio de orientación psicopedagógico del tribunal, del delegado de libertad vigilada y del centro de reforma o de educación, además de las diligencias previas y el *acuerdo* del tribunal). Estos *informes* proporcionaron una información valiosa e imprescindible para conocer el funcionamiento de la justicia de menores: cómo se acuerdan y ejecutan las medidas punitivas.

A partir de estos primeros *expedientes* se elaboró una ficha ad hoc para organizar la recogida de información, a modo de *cuestionario* a rellenar con los datos extraídos de cada uno de los *expedientes*. Su *validez* fue previamente com-

probada en una *muestra* de 60 *expedientes* extraídos aleatoriamente del total de expedientes a analizar. Incluía información relativa al:

- *Menor*: sus características sociodemográficas y delictivas. La aplicación conjunta de técnicas analíticas multivariantes (de *conglomerados*, *discriminante* y *factorial confirmatorio*) contribuyó a diferenciar tres tipos de menores, rompiendo el común estereotipo unívoco del menor delincuente.
- *Proceso de aplicación de la norma legal*, desde la instrucción del procedimiento penal hasta el *acuerdo* de medida adoptada por el tribunal. Se contrastó la información obtenida del expediente (lo que “es”), con lo dicho al respecto en la normativa legal vigente (lo que “debería ser”). En concreto, se analizaron los diferentes criterios seguidos por el tribunal para adjudicar *tratamientos* a los menores (a partir de la instrucción del procedimiento y del *acuerdo* tomado finalmente por el juez). El análisis de *regresión logística* en los datos registrados permitió obtener las variables que se hallaban más relacionadas con la probabilidad de imponer al menor los distintos *tratamientos* (medidas leves, libertad vigilada e internamiento).
- La *aplicación del tratamiento*. La *evaluación del proceso* concluyó con la descripción de cómo se ejecuta realmente el *programa*: si los diferentes *tratamientos* acordados por el tribunal se aplicaron en conformidad con las normas establecidas. Para ello de nuevo se contrastó la información extraída del *expediente* con las disposiciones expresas en el texto legal (LTTM de 1948).

3. *Evaluación del impacto*. Toda *evaluación del impacto* debe, primero, definir claramente cuáles son los criterios de éxito y de fracaso del *programa*; y, segundo, diferenciar aquellos que se deben al programa de los causados por otros factores (McQuay, 1979). Definir los criterios de éxito de la intervención judicial no resulta difícil. Habrá éxito siempre y cuando el *tratamiento* proporcionado consiga prevenir la *reincidencia* en conducta delictiva, logrando la reinserción del menor en la sociedad. El problema surge en su operacionalización. La *no reincidencia* hay que medirla tanto durante el tiempo de aplicación del *tratamiento* como al menos dos años después de su conclusión. Del *expediente* del menor se extrajo el número de detenciones policiales registradas antes y durante el período de *tratamiento*. Pero, para obtener información de detenciones posteriores a la actuación del Tribunal, había que acudir a otras instancias: la Dirección General de la Policía y la Dirección General de Instituciones Penitenciarias.

La Dirección General de la Policía primeramente rehusó facilitar la información requerida argumentando impedimentos legales: la anticonstitucionalidad de proporcionar información de detenciones policiales de personas concretas. A través de *redes interpersonales* la autora conoció al comisario responsable del Centro de Datos de la Policía. Después de más de tres meses de gestiones, al final obtuvo la información solicitada: el número de detenciones y de reclamaciones judiciales o policiales posteriores al cierre del *expediente* del menor, fecha y motivo del cierre, para cada uno de los 715 menores. El pleno anonimato de los datos identificativos de estos menores tuvo que garantizarse por escrito.

Después se personó en la Dirección General de Instituciones Penitenciarias, que mostró aún más reticencias a dar información. A los impedimentos legales se

sumaron dificultades informáticas (de personal, medios). Tras numerosas gestiones, en las que la Dirección General de Protección Jurídica del Menor tuvo incluso que mediar, y un compás de espera de casi un año, al final se consiguió saber qué jóvenes estaban o habían estado en alguno de los establecimientos penitenciarios de la Comunidad de Madrid.

Con la información ya recabada, se procede a su análisis. El objetivo era conocer el peso de la población reincidente en el conjunto estudiado y las variables que determinaban la probabilidad de reincidencia (mediante análisis de *regresión logística*). Para comprobar la eficacia de dos de los *tratamientos* aplicados (libertad vigilada e internamiento), primero se dividió a la población en tres grupos, de acuerdo con el *tratamiento* que recibieron: sólo libertad vigilada, sólo internamiento, y ambos *tratamientos*. Éstos no constituían grupos homogéneos (aunque pudieran asemejarse en algunos rasgos), ni habían sido formados a priori, de forma aleatoria. Los sujetos pasaron a uno u otro *tratamiento* en función de sus características sociodemográficas y delictivas.

Se calculó la proporción de individuos que reincidieron en conducta delictiva (para cada una de las variables consideradas). Después se aplicó el análisis de *regresión logística* para obtener las variables que más determinan la probabilidad de reincidir. En esta fase de la investigación interesaba conocer cómo incidía cada una de las circunstancias personales, sociofamiliares y legales del joven en la *reincidencia*; además de aspectos relacionados con la aplicación del *tratamiento* (su duración, el número de diligencias policiales y de denuncias previas, durante y posteriores al *tratamiento*, la adaptación del menor y el motivo del cese). Sus efectos se controlaron por el tipo de *tratamiento* que se aplicó.

A partir de la información recabada en cada una de las fases de la investigación, se procedió a la *interpretación* de los resultados del *programa*. Éstos resultaron ser bastante negativos. Dos de cada tres menores proseguían su *carrera delictiva*, después de haber abandonado la jurisdicción de menores. La información obtenida durante la *evaluación del proceso* ayudó a la búsqueda tanto de las *causas* del elevado porcentaje de reincidencia como de las *soluciones* (la plena modificación del *programa* o conjunto de actuaciones de los tribunales tutelares de menores en España).

- b) *Datos publicados* por organismos públicos y privados: *estadísticas e informes*, que editan como parte principal (o complementaria) de su actividad. Incluye los informes y “estadísticas oficiales” que elaboran las agencias oficiales o institutos estadísticos de los diferentes países para describir y cuantificar su población y actividad.

Por ejemplo:

- Las estadísticas elaboradas por *Eurostat* (Oficina Estadística de la Unión Europea) sobre economía y finanzas, población y condiciones sociales, industria, servicios y comercio, agricultura, pesca y transporte,

medio ambiente y energía, y ciencia y tecnología en los países de la Unión Europea; además de sus informes como *Europe in Figures - Eurostat Yearbook* o *Income and Living Conditions in Europe*, que ofrece cifras sobre el empleo, desigualdad de renta, pobreza, vivienda, salud, educación y exclusión social.

- También están las estadísticas e informes elaborados por la Organización de Cooperación y Desarrollo Económico (OCDE), como *OECD Statistical Compendium* (un compendio de sus principales bases de datos estadísticos); las realizadas por Naciones Unidas, como el *Statistical Yearbook* o el *Demographic Yearbook*, o las estadísticas específicas de la UNESCO, entre otras organizaciones dependientes de la ONU.
- A nivel nacional (en España), las estadísticas que elabora el *Instituto Nacional de Estadística* (INE): el centro responsable de la realización del *Censo de Población* (desde 1857). En 2011 deja de recopilar información “casa por casa” y combina, por vez primera, el uso de registros administrativos con el trabajo de campo. Incluye un *Censo de Edificios* y una gran encuesta al 12% de la población (unos 5,7 millones de personas, 3 millones de viviendas), con el objetivo de ahorrar circa 300 millones de euros. Además elabora el *Anuario estadístico de España* (desde 1858), el *Nomenclátor* (desde 1858), el *Movimiento natural de la población* (desde 1858), el *Censo de edificios y viviendas* (desde 1950), las *Estadísticas judiciales* (desde 1959), la *Encuesta de población activa* (desde 1964), la *Proyección de la población española* (desde 1978), la *Encuesta de migraciones* (desde 1980), y la *Encuesta continua de presupuestos familiares* (desde 1985), entre otras (que se detallan en su página web).
- Las estadísticas que confeccionan los diferentes ministerios (caso de las estadísticas sobre el mercado de trabajo, inmigración y emigración del anterior Ministerio de Trabajo e Inmigración), los bancos centrales de los distintos países, los departamentos de estadística de ámbitos regionales (comunidades autónomas) y locales (ayuntamientos), etc. Sus estadísticas e informes son, asimismo, de fácil acceso a través de sus correspondientes páginas webs.
- Los datos publicados por organismos o *entidades privadas* como, por ejemplo, los servicios de estudios de entidades bancarias (como Banesto, BBVA, la Caixa o la Confederación de Cajas de Ahorros...) o los centros de investigación social. Entre estos está ASEP/JDS (fundado en 1982 por el catedrático de Sociología Juan Díez Nicolás), en cuyo *banco de datos* se hallan informatizadas todas las encuestas que realiza desde 1986 y que están accesibles, de modo interactivo, a través de la Red. Su *banco de datos* también incluye los *Latinobarómetros* (un estudio de opinión pública realizado en 18 países de América Latina), los *Estudios de valores* (datos recogidos por *World Values Survey*, WVS, y *European Values Study*, EVS, desde 1981), entre otros.

Para el *análisis secundario* de datos de encuesta (reanálisis de encuestas realizadas por otros) un gran potencial lo ofrece el *Banco de datos* del Centro de Investigaciones Sociológicas (CIS), dependiente del Ministerio de la Presidencia. Reúne la información de todas las encuestas y estudios cualitativos que el centro ha llevado a cabo desde su creación en 1963 (como Instituto de la Opinión Pública). Todos los datos almacenados en este repositorio son de acceso público. A partir de enero de 2009, ofrece gratuitamente la descarga de los ficheros de microdatos de los estudios realizados por el centro durante los últimos años.

En el ámbito europeo está el *Banco de datos* de la Comisión Europea. Incluye todos los *eurobarómetros* realizados desde 1973, con el objetivo de seguir la evolución de la opinión pública en todos los países miembros de la Unión Europea. También destaca el *Economic and Social Research Council* (ESRC) de la Universidad de Essex (Inglaterra) o los *Steinwetz Archives* en Holanda.

En EE UU, los más referenciados son *Inter-University Consortium for Political and Social Research* (ICPSR) de la Universidad de Michigan (el mayor archivo de datos en ciencias sociales); *National Opinion Research Center* (NORC), de la Universidad de Chicago; *Survey Research Center*, de la Universidad de California-Berkeley; o *The Institute for Research in Social Science*, de la Universidad de North Carolina-Chapel Hill. De la serie de encuestas que incluyen, la *General Social Survey* (realizada por NORC todos los años desde 1972) es la más ampliamente utilizada, por los investigadores sociales, para el análisis de datos *secundarios*. Al igual que los barómetros de opinión del CIS o los eurobarómetros de la Comisión Europea, la encuesta repite unas mismas preguntas y tópicos todos los años; a éstas se añaden tópicos diferentes en años específicos.

Además está el archivo de datos online del *International Social Survey Program* o el correspondiente a la *European Social Survey*, que cubre a una treintena de países (principalmente europeos).

Por último, reiterar que el *análisis “secundario”* es el análisis posterior de información ya existente. El foco está en el análisis más que en la recogida de datos. “Tal análisis puede estar relacionado con el propósito original para el que los datos se recogieron, o puede dirigirse a un asunto bastante diferente del que instó el esfuerzo de reunión de los datos originales. Puede implicar la integración de distintas fuentes o un reanálisis de los datos de una fuente única” (Stewart, 1984: 11). Facilita el análisis comparativo y el de tendencias, a partir de los datos disponibles para un período amplio de tiempo. De acuerdo con Hakim (1994: 23), “es probable que siga siendo la aproximación más corriente para realizar estudios comparativos internacionales, especialmente para estudios que buscan cubrir un número elevado de países o tendencias a lo largo del tiempo”. De hecho, una proporción importante de “investigación económica se basa en el análisis secundario de se-

ries temporales de nivel macro, que consisten en un gran número de indicadores estadísticos nacionales y medidas cotejadas desde una gran variedad de encuestas oficiales y series estadísticas” (Hakim, 1994: 22). Lo que es extensible a otras áreas de conocimiento.

Su auge es creciente en los últimos años. A ello contribuye su relativo bajo coste y que posibilita la comparación entre ámbitos locales, nacionales e internacionales, a lo largo del tiempo: el seguimiento comparativo longitudinal. Neuman (2000) añade su potencialidad para la replicación y el permitir abordar temas no pensados por los investigadores que originaron la información.

La recogida de datos a gran escala es cara y difícil, no siendo accesible a la mayoría de los investigadores. Afortunadamente en los últimos años se ha ampliado y mejorado la sensibilidad hacia la necesidad de “archivar” la información y hacerla accesible a la población interesada en su conocimiento y análisis. Esta fase de apertura de *archivos* y *bancos de datos* sin duda se ha acelerado con la extensión de Internet no sólo en los centros universitarios y de investigación, sino también en los hogares.

EJEMPLO DE ANÁLISIS SECUNDARIO

- A) Cea D’Ancona, M.^a Á. (2007: *La deriva del cambio familiar. Hacia formas de convivencia más abiertas y democráticas*. CIS. Monografía 241).

El objetivo de la investigación era comprobar si, con el nuevo siglo, los cambios antes detectados hacia una mayor permisividad y democratización en las convivencias familiares se mantienen, se intensifican o, por el contrario, se produce un retroceso o menor apoyo social ante la pluralidad familiar. Para ello se contrastaron los datos de opinión más recientes con datos de *archivos estadísticos*. Se diseñó una investigación *triangular*, que conjugaba el *análisis secundario* de datos de encuesta con el correspondiente a *documentos* y *archivos estadísticos*. Es un estudio *comparativo* (los cambios habidos en España se compararon con los registrados en los demás países de la Unión Europea); *longitudinal de tendencias* (se comprobó en qué medida se ha avanzado, estancado o retrocedido respecto a fechas anteriores); y *explicativo* (se indagó en los factores que pueden haber coadyuvado a los cambios de los modelos familiares).

A través de los datos de *encuesta* se calibró el alcance de los cambios familiares y los *valores* que predominaban en las relaciones familiares en cada momento temporal. Como se precisaba comprobar si los modelos ideales y las normas sobre la organización de la vida familiar coincidían con los comportamientos efectivos, los datos de encuesta se cotejaron (o contrastaron) con los datos correspondientes a *estadísticas oficiales* para ver en qué medida describían una misma “realidad”. Las encuestas que son objeto de explotación estadística exhaustiva procedían del *banco de datos* del CIS. En concreto, fueron diez las encuestas que se analizaron. Éstas se especifican en el cuadro 4.1, donde aparecen ordenadas según los objetivos de investigación a cubrir.

CUADRO 4.1. Encuestas del Centro de Investigaciones Sociológicas (CIS) analizadas

Número estudio	Título	Objetivo de investigación
2.568	Barómetro junio 2004.	1.º Conceptualización de la familia y permisibilidad ante formas alternativas de convivencia en familia.
2.248	Barómetro mayo 1997.	
2.087	Barómetro abril 1994.	
2.556	Barómetro febrero 2004.	2.º Estructura de autoridad de la familia (en la pareja y hacia los hijos): toma de decisiones y distribución de derechos y deberes.
2.529	<i>Familia y género (International Social Survey Programme)</i> : junio 2003.	
2.113	<i>Family and changing gender roles</i> : septiembre 1994.	3.º Relaciones intergeneracionales: comunicación entre padres e hijos.
2.536	Sondeo sobre la juventud española 2003 (3.ª oleada): septiembre 2003.	Incluyen indicadores clasificados en los objetivos 2 y 3. El sondeo de 1997 participa además en dimensiones cubiertas en el objetivo 1.
2.469	Sondeo sobre la juventud española 2002 (3.ª oleada): octubre 2002.	
2.262	Juventud y entorno familiar: octubre 1997.	

Los datos de estas encuestas se compararon con los obtenidos en otros estudios a escala nacional y europea (*eurobarómetros*, principalmente), para unas mismas fechas e indicadores. La finalidad era doble: analizar en qué medida reflejaban un mismo estado de opinión, y situar a España en el contexto de otros países de similar nivel de desarrollo socioeconómico y político.

Para la contextualización histórico-comparativa fue asimismo imprescindible el acceso a los *archivos estadísticos*. Los datos estadísticos permitieron cifrar la magnitud de las transformaciones familiares, además de corroborar los datos de opinión sobre cada aspecto concreto. A escala nacional, fundamentalmente se analizaron datos estadísticos elaborados por el Instituto Nacional de Estadística (INE); para el contexto europeo, los proporcionados por la Oficina Estadística de la Unión Europea (*Eurostat*). A través de ellos se indagó en los cambios demográficos y en los laborales, por su efecto directo en la vida familiar.

El uso combinado de diversas técnicas analíticas multivariantes (*factorial, conglomerados y discriminante*) en los datos de encuesta permitió la elaboración de tipos de modelos de familia, a partir de los indicadores analizados en cada apartado. La existencia de unos mismos indicadores en encuestas temporalmente espaciadas permitió trazar la evolución de los diferentes modelos de familia, además de cuantificar su presencia en contextos sociotemporales diversos.

B) El mismo diseño de investigación *secundaria* se aplica en los estudios, que la autora realiza sobre inmigración, racismo y xenofobia: desde la monografía publicada en 2004 (*La activación de la xenofobia en España. ¿Qué miden las encuestas?* CIS/Siglo XXI. Colección *Monografías* n.º 210) hasta la serie de informes sobre la *Evolución del Racismo y la Xenofobia en España* que desde 2007 efectúa para el Observatorio Español de Racismo y Xenofobia (OBERAXE).

En dichas investigaciones se aplica la estrategia metodológica de la *triangulación*. Principalmente consiste en conjugar:

- *Análisis secundario de datos de encuesta*: los datos de opinión procedentes de las encuestas levantadas por el CIS (a partir de 2007 encargadas por OBERAXE) se complementaron con los obtenidos por otros centros de investigación en España (ASEP, IESA) y Europa (*Eurobarómetros*, *la Encuesta Social Europea*), para unos mismos indicadores y fechas de encuesta.
 - *Datos de archivos estadísticos y materiales cualitativos* (correspondientes a entrevistas y grupos de discusión de investigaciones propias). Se analizaron datos estadísticos proporcionados por el INE, Eurostat, el Ministerio del Interior y el Ministerio de Trabajo e Inmigración, que cifran la “realidad” de la inmigración.
 - *Revisión bibliográfica* (de investigaciones teóricas y empíricas) de temas relacionados con el objeto de estudio: *inmigración*, *racismo* y *xenofobia*. Esto último necesario para configurar el *marco teórico* de la investigación e interpretar los datos de encuesta. Como también lo es el contraste con *datos estadísticos*. Estos últimos permiten indagar en la imagen estereotípica de la inmigración, en su congruencia o disonancia con la realidad.
 - *Contextualización mediática*: el análisis de las noticias publicadas en la *prensa* en fechas previas a la realización del trabajo de campo de las *encuestas* que son objeto de análisis. Ello se debe al papel demostrado que los medios de comunicación ejercen en la configuración de la imagen colectiva de la inmigración, en la génesis y extensión de la *xenofobia*.
-

4.2.3. Meta-análisis

Se define como “el análisis de los análisis (el análisis estadístico de los hallazgos de muchos análisis individuales)” (Glass *et al.*, 1987: 21). “No es un método de investigación o una fuente de datos”, sino que “proporciona un medio de resumir los resultados de numerosos estudios cuantitativos sobre un dominio particular” (Bryman, 1995: 228).

El término fue aplicado por primera vez por Gene V. Glass en 1976; si bien en 1904 Karl Pearson ya lo promovió como propuesta para superar el problema de la reducida significatividad estadística de estudios con tamaños muestrales pequeños: analizando los resultados de un grupo de estudios similares se puede alcanzar una valoración más exacta de los efectos.

El objetivo principal es la obtención de un cuadro resumen que sintetice los resultados cuantitativos alcanzados en diferentes investigaciones. Cada estudio revisado constituye una unidad de la *muestra* de estudios (relevantes), extraída del total existente. En ella se aplican distintas *técnicas* de análisis estadístico (*univariable*, *bivariable* o *multivariable*), que permitan comprobar la *significatividad* de los hallazgos de los estudios que se analizan. Los *hallazgos* se convierten en la variable de-

pendiente (se observa su variación de un estudio a otro), y las características (sustantivas y metodológicas) de los estudios en las *variables independientes*. Se determina la magnitud de los *efectos* y su significatividad.

Las fases del *meta-análisis* se asemejan a las de una investigación “*primaria*” (delimitación de objetivos, hipótesis...), con la salvedad de que la unidad de observación son los estudios revisados. Habrá que indicar:

- Los *ámbitos* para la búsqueda de las investigaciones.
- Los *criterios* de inclusión y exclusión (criterios de calidad que incluyan las diferentes variantes de validez interna y externa).
- El procedimiento para resumir los datos.

Después se analizan y codifican los estudios seleccionados de acuerdo con determinadas características. Para ello se:

- Determina qué datos, características o información interesa de cada estudio.
- Elaboran cuadros que sintetizen la información.
- Calcula la magnitud del efecto de cada estudio y la relación entre determinadas características y el efecto.

En los *cuadros* suele aparecer: autores, título del estudio, fecha, características de la fuente y del estudio, tipo de muestra, variable dependiente, diseño y análisis estadístico (*t* de *Student*, *Anova*, *Chi-cuadrado*...), conclusiones finales. Se calcula el tamaño del *efecto* en cada estudio y de la *media* de todos.

De acuerdo con Glass *et al.* (1987: 69), la *cuantificación* suele implicar “la medición de aspectos métricos (*por ejemplo*, ¿en qué año se realizó el estudio?, ¿cuál es el tamaño de la *muestra* en la que se obtiene el *coeficiente de correlación* r_{xy} ?) y su función nominal o de codificación (*por ejemplo*, ¿fueron las diferencias iniciales corregidas por el *análisis de la covarianza*? Sí = 1, No = 2)”. A ello se suma la aplicación de *tests de significatividad* (como la *t* de *Student* o la χ^2).

La solidez de las conclusiones del *meta-análisis* dependerá de la calidad de los estudios revisados: si “representan” realmente a la totalidad de estudios existentes (investigaciones publicadas y no publicadas). También, de cómo se haya procedido en el análisis de cada uno de ellos. La crítica común al *meta-análisis* es que tiende a centrarse en el resultado final (el tamaño del efecto), y descuida las características de los estudios. Su *validez* dependerá de que la revisión sea completa y de que los estudios seleccionados constituyan una *muestra* representativa.

EJEMPLO DE META-ANÁLISIS

El *meta-análisis* realizado por Pedro J. Moreno Gil, F. Xavier Méndez Carrillo y Julio Sánchez Meca (2000: “Eficacia de los tratamientos cognitivo-conductuales para la fobia social: una revisión meta-analítica”. *Psicothema*, vol. 12, n.º 3: 346-352).

El objetivo del *meta-análisis* era estimar la eficacia de las técnicas de exposición, el entrenamiento en habilidades sociales, las técnicas de reestructuración cognitiva y sus combinaciones en pacientes que sufren fobia social; además de determinar las variables que moderan la eficacia terapéutica en este trastorno fóbico.

En los *meta-análisis* distinguieron entre “informe de investigación” y “estudio”:

- *Informe de investigación*: el soporte empleado para informar sobre los resultados de una o más comparaciones de grupos relativos a la eficacia de tratamientos psicológicos para la fobia social.
- *Estudio*: la comparación de un grupo que recibe un tratamiento psicológico con un grupo de control. Cuando no se comparaba un grupo tratado con otro no tratado, pero se evaluaba la ganancia terapéutica comparando una medida *postest* con una medida *pretest* para dicho grupo que recibe tratamiento, se considera igualmente que constituye un *estudio* independiente.

Con el objeto de homogeneizar la muestra empírica, y aplicar adecuadamente la tecnología *meta-analítica*, los autores establecieron los siguientes *criterios conceptuales y metodológicos* para incluir y excluir informes de investigación:

- a) El *informe* de la investigación debía contrastar la eficacia de técnicas de exposición, técnicas de reestructuración cognitiva, entrenamiento en habilidades sociales o alguna combinación de estos tratamientos.
- b) Los sujetos tratados debían recibir un diagnóstico de fobia social.
- c) El concepto de *fobia social* que se empleaba en dicho informe de la investigación debía ajustarse, de modo explícito, a las definiciones incluidas en los sistemas nosológicos CIE (Organización Mundial de la Salud) o DSM (*American Psychiatric Association*).
- d) La fecha de publicación (o realización del estudio, si éste no había sido publicado) debía estar comprendida en el intervalo 1980-1997, ambos inclusive. El límite inferior de dicho intervalo se determinó por la fecha de publicación de la tercera edición del *Diagnostic and Statistical Manual of Mental Disorders*.
- e) El *diseño de la investigación* debía ser un diseño de grupo, con al menos 5 sujetos experimentales.
- f) El *informe* de la investigación debía aportar suficientes datos para estimar los tamaños del efecto necesarios.
- g) Se excluyeron los *informes* de investigación que no presentaban datos de *pretest*.
- h) Se excluyeron también los *informes* que empleaban la metodología de *diseño de caso único*, ya que no es posible integrar cuantitativamente los resultados de diseños de grupo con los diseños de caso único.

Para minimizar sesgos de selección, se emplearon varios *procesos de búsqueda de informes de investigación*:

- *Consultas informatizadas*: utilizaron las bases de datos PsycLIT, MEDLINE y PSI-CODOC. Los descriptores empleados fueron: “fobia social”, “ansiedad social”, “tratamiento”, “terapia”, “social phobia”, “social anxiety”, “treatment”, “therapy”.
- *Revisión de artículos y monografías* de autores relevantes en el área.
- *Revisión de referencias bibliográficas* de los informes de investigación ya localizados como fuente de estudios primarios previos.
- *Petición de trabajos a expertos*, publicados y no publicados.

La *revisión bibliográfica* les permitió localizar 25 *informes de investigación* que cumplan con los criterios de selección, arrojando un total de 39 *estudios*.

Codificación de estudios: para definir operacionalmente las variables que debían ser codificadas. Confeccionaron un manual con pautas específicas y un protocolo de codificación. La calidad del proceso de codificación fue estudiada mediante la selección de una *muestra* aleatoria del 33% del total de *informes* localizados, que fue codificada por dos codificadores independientes.

Las *variables* cuyo posible efecto moderador se contrastó fueron clasificadas en tres grandes apartados:

1. *Variables sustantivas*: son propias del ámbito de investigación objeto del *meta-análisis*. Incluyen variables de *tratamiento* (tipo de tratamiento psicológico aplicado, género de los profesionales, experiencia de los profesionales, tareas para casa del tratamiento, programa de seguimiento del tratamiento, utilización de agentes externos al grupo terapéutico, contrato terapéutico, modalidad de entrenamiento, número de terapeutas, número de sesiones de tratamiento, duración del tratamiento, intensidad del tratamiento, absentismo), de *sujeto* (edad, género, nivel socioeconómico, nivel educativo, sistema de diagnóstico empleado, subtipo de fobia social, duración media de la fobia social, presencia de tratamientos previos, tipo de tratamientos psicológicos y farmacológicos previos), y de *contexto* (continente, país).
2. *Variables metodológicas*: los aspectos relacionados con el diseño y la metodología de la investigación (modo de reclutamiento de los sujetos, medidas para preservar la integridad del tratamiento, tipo de diseño, asignación aleatoria a los grupos, tipo de grupo de control, tamaño muestral total, tamaño muestral del grupo de tratamiento, tamaño muestral del grupo de control, mortalidad experimental, calidad del diseño, meses de seguimiento, número de variables dependientes).
3. *Variables extrínsecas*: no se relacionan ni con la metodología ni con el ámbito de investigación. Aunque se trata de variables que no deben afectar a los resultados, en ocasiones pueden ser relevantes (formación técnica del primer investigador, forma de presentación del estudio, rango de difusión, fecha de presentación del estudio).

Cálculo del tamaño del efecto: el índice del tamaño del efecto (TE) que emplearon es la diferencia media tipificada d , atendiendo a las siguientes definiciones según la naturaleza del diseño que presenta el estudio que se evalúa:

- Para los *diseños intragrupo* (*pretest-postest*), d se definió como la diferencia entre la media del *pretest* y la media del *postest* dividida por la desviación típica intragrupo conjunta (o en su defecto, el promedio de las desviaciones típicas del *pretest* y del *postest*).

- Para los *diseños intergrupos* (tanto experimentales como cuasiexperimentales), con medidas *pretest* y *posttest*, d se definió como $d = dE - dC$, siendo dE y dC las diferencias medias tipificadas entre el pretest y el posttest de los grupos *experimental* y de *control*, respectivamente. Un valor d mayor de cero indica un efecto beneficioso para los sujetos del grupo tratado; un valor d menor de cero indica un efecto perjudicial.

Con el objetivo de optimizar la comparación entre los tratamientos meta-analizados, maximizaron la homogeneidad de los TE. Se seleccionaron sólo las variables dependientes directamente relacionadas con la definición de *fobia social*; es decir, miedo, evitación o malestar en situaciones sociales, que implica deterioro de la adaptación social, laboral o académica, y se excluyen las restantes (rasgo de ansiedad, ánimo depresivo, autoestima, locus de control, creencias irracionales, etc.).

Para cada estudio se calculó un máximo de dos TE: uno para la evaluación *posttest* y otro para la evaluación del seguimiento más largo. Para cada momento temporal se promediaron los valores d derivados de las variables dependientes que se ajustaban a los criterios establecidos.

Técnicas de análisis estadístico: se calculó un intervalo de confianza en torno al TE medio para estimar la magnitud del efecto en la población y si dicho efecto era significativamente distinto de cero (efectividad nula). Además se aplicó una prueba χ^2 de homogeneidad de todos los tamaños del efecto en torno al TE medio. Dado que la prueba de homogeneidad puede presentar baja potencia cuando se aplica a un grupo reducido de estudios, como es el caso, decidieron comprobar el influjo de variables que teóricamente podrían estar moderando los resultados, aunque no resultara significativa dicha prueba de homogeneidad. Siguiendo el enfoque meta-analítico de Hedges y Olkin para las variables cualitativas, se realizaron análisis de varianza ponderados por la inversa de la varianza de cada TE. Para las variables cuantitativas se aplicaron análisis de regresión simple ponderados por la inversa de la varianza de cada TE.

Fiabilidad de la codificación: la fiabilidad intercodificadores para las variables moderadoras alcanzó los valores que garantizan la *fiabilidad* de la codificación (al menos un valor de “0,80” para la tasa de acuerdo y para el coeficiente de *correlación de Pearson*, y al menos un valor de “0,70” para el coeficiente *Kappa de Cohen* y la *correlación intraclase*). La *fiabilidad* de los cálculos del tamaño del efecto (correlación intraclase y correlación de Pearson) superó el valor de “0,98” para el posttest y el seguimiento, lo que indica un excelente consenso en la selección de variables dependientes a incluir en el cálculo del tamaño del efecto.

Estudio del sesgo de publicación: la totalidad de los estudios incluidos en el presente *meta-análisis* estaban contenidos en informes de investigación publicados en revistas especializadas. Por lo que una posible amenaza a la *validez* de los resultados obtenidos se encontró en un eventual *sesgo de publicación* (que los editores aplicaban un trato dispar a los estudios presentados para su publicación en función de la significación estadística informada). Calcularon el “índice de tolerancia a los resultados nulos”, según el cual debía existir más de 180 estudios no publicados (y no recuperados por el meta-analista) archivados en las editoriales de las revistas para que los resultados del *meta-análisis* queden anulados. De lo que se derivó que era poco probable que el *sesgo de publicación* estuviese afectando a los resultados del *meta-análisis*.

Del *meta-análisis* se concluyó que los tratamientos cognitivo-conductuales revisados resultaban claramente efectivos para la *fobia social*. El tamaño del efecto medio ponderado (por la inversa de la varianza) fue de “0,769” en el *posttest* y “0,953” en el seguimiento, hallándose sus valores paramétricos lejos del valor nulo. Además, el valor medio obtenido en el *meta-análisis* se aproximó a una magnitud alta en el *posttest* ($d = 0,80$), superándose en el seguimiento. La prueba de homogeneidad no resultó estadísticamente significativa [*posttest*: $QT(38) = 19,163$, $p > .05$; *seguimiento*: $QT(33) = 25,318$, $p > .05$]. Por lo que dedujo que los distintos estudios podían considerarse homogéneos entre sí, pese a derivarse de estudios que diferían en un elevado número de características de tipo sustantivo y metodológico, resultando sus valores paramétricos representativos de los hallados en los estudios meta-analizados.

4.2.4. Análisis de contenido cuantitativo

En el análisis de *documentos* (y, en general, de cualquiera de las modalidades de comunicación verbal y visual) adquiere un destacado protagonismo el *análisis de contenido*: una técnica para reunir y analizar el contenido de un *texto*, que se considera “no reactiva”. “El proceso de poner palabras, mensajes, o símbolos en un texto para comunicar a un lector o receptor ocurre sin influencia del investigador que analiza su contenido” (Neuman, 2000: 293). El *contenido* se refiere a palabras, significados, dibujos, símbolos, temas o cualquier mensaje que pueda ser comunicado. El *texto* es algo escrito, visual o hablado, que sirve como un medio para la comunicación. Incluye libros, artículos de prensa, anuncios, discursos, documentos oficiales, transcripciones de entrevistas, películas, vídeos, fotografías... En 1910 Max Weber ya lo sugirió para el estudio de las noticias aparecidas en la prensa. El interés está en analizar el “contenido” *manifiesto* y *latente* de la información verbal o visual que sea objeto de estudio.

Tesch (1992) diferencia dos tipos genéricos de *análisis de contenido*: el *clásico* y el *etnográfico*. El primero se ajusta a la definición de Berelson (1952) como la descripción *objetiva, sistemática, y cuantitativa* del *contenido manifiesto* de la comunicación; mientras que el segundo (el *etnográfico*), responde a una redefinición desde posturas *cualitativas* que ponen el acento en el *análisis reflexivo de documentos*. Este segundo tipo se dirige, no tanto a la *descripción*, como a la *comprensión* de los *significados latentes*, y a la *verificación* de relaciones teóricas. Aquí únicamente se va a ofrecer una visión sintética del *análisis de contenido clásico* o *cuantitativo*, que utiliza técnicas de análisis cuantitativas. Para su complementación con el *análisis de contenido etnográfico*, se remite a manuales de metodología cualitativa.

La realización de un *análisis de contenido* cuantitativo comprende las siguientes fases:

1. *Preanálisis* o fase de organización, de *operacionalización* de las ideas originarias de la investigación. Incluye:

- La especificación de los *objetivos* de la investigación y la formulación de las *hipótesis*.
 - Elaboración de los *indicadores* en que se apoyará la interpretación de los datos. Estos se obtendrán a partir de las *definiciones operativas* de los *conceptos teóricos* básicos en la investigación.
 - Elección de los *documentos* a analizar: la *muestra* del estudio. Ha de delimitarse el “*corpus*” o material a analizar. Incluye la *población* de documentos y la concreción del procedimiento de *muestreo* a seguir para su selección. Lo habitual, salvo que el *universo* sea pequeño, es extraer una *muestra* representativa aplicando algún procedimiento aleatorio (como se describe en el capítulo 7). La selección *muestral* comienza con la elección de la *f fuente de comunicación* (periódicos, revistas, libros, canciones, imágenes); prosigue con la extracción de *documentos* y, en caso de que éstos sean muy extensos, de *partes* de estos documentos.
2. *Explotación del material*. Incluye los criterios para la extracción de la información “contenida”, de acuerdo con los *objetivos* de la investigación y las características de los documentos. Comprende:
- Elección de las *unidades de codificación*: palabras, expresiones, frases, temas, imágenes... (depende de qué se analice). Se cuantificará la *frecuencia* (e *intensidad*) de aparición en el documento. Además pueden efectuarse enumeraciones de carácter temporal (cuándo) y espacial (dónde aparecen). También es importante la identificación de las *unidades de contexto*, ya que una misma palabra puede adoptar significados diferentes, dependiendo del *contexto* donde se ubique. La contextualización ayudará a la *interpretación* del significado de las *unidades de codificación*.
 Si la *unidad de registro* fuera, por ejemplo, la *palabra*, la *unidad de contexto* sería la *frase*; si fuese el *tema*, el *párrafo*; si fuera el *párrafo*, el *documento* entero.
 - Determinación de las *categorías* para la *codificación* del “contenido” (a partir de las *hipótesis* de investigación), siempre y cuando se realice un *análisis de contenido categorial* o temático. Las *categorías* proporcionan una denominación genérica a unidades de contenido similar, contribuyendo a su clasificación. Antes de su aplicación definitiva, conviene comprobar el *esquema de codificación* en una pequeña *muestra* (de la misma población de documentos). Ayudará a precisar su adecuación para los objetivos de la investigación.
 - *Registro y cuantificación* de los *datos*, de acuerdo con los *indicadores* elegidos.
3. *Tratamiento e interpretación de los resultados*. Como en la experimentación o la investigación mediante encuesta, a los datos reunidos se les da tratamiento

estadístico (descrito en el capítulo 8). Se comienza con análisis *univariable* y, si las características de los datos lo permiten, puede concluirse con la aplicación de alguna técnica de análisis *multivariable*. Una de las *técnicas multivariabiles* más aplicadas en el *análisis de contenido* es el *análisis factorial*. En éste, las *variables observadas* suelen ser las *categorías* diferenciadas; y las *latentes* hacen referencia a *temas* genéricos (“latentes”) en el texto.

Los resultados numéricos se *validan e interpretan* en relación con el *tema* al que hacen referencia. Los resultados pueden corroborar las *hipótesis* de la investigación o, por el contrario, contradecirlas. Lo que puede provocar la realización de una nueva investigación.

El *análisis de contenido*, como cualquier proceso de medición, ha de satisfacer requisitos de *validez* y *fiabilidad*. La *validez* indica la *adecuación* del sistema de clasificación seguido para medir los *conceptos* teóricos. La *fiabilidad* expresa la *consistencia* del procedimiento de medición. Respecto a la *validez* en el *análisis de contenido* (López Aranguren, 1989/ 2000; Krippendorff, 1990; Weber, 1994), se distinguen variantes:

- a) *Validez pragmática* o de utilidad de los resultados del análisis para el estudio del tema que se investiga.
- b) *De constructo*: correspondencia entre la medida aplicada con alguna otra utilizada del mismo *constructo*.
- c) *De contenido*: grado al que los *indicadores* cubren la variedad de significados incluidos en el *concepto*.
- d) *Predictiva*: grado de acuerdo o correspondencia entre las predicciones obtenidas y los hechos observados.
- e) *Semántica*: grado al que las *unidades de codificación* clasificadas juntas poseen connotaciones similares. En el *análisis de contenido* “se logra una alta validez semántica, cuando la semántica del lenguaje de los datos se corresponde con la fuente, el receptor o cualquier otro contexto respecto al que se examinan dichos datos” (Krippendorff, 1990: 231).

Por lo que, los límites a la *inferencia* se hallarán en la ambigüedad de los significados de las *palabras* y de las *reglas de codificación*. Razón por la cual se recomienda su explícita delimitación.

Respecto a la *fiabilidad* (o consistencia del proceso de medición), han de arbitrase unas reglas de *codificación* que permitan que diferentes codificadores lleguen a unos mismos resultados (independientemente unos de otros). Esta meta es más fácil de alcanzar cuando la *codificación* se hace mediante ordenador y las *reglas de codificación* están bien especificadas. Se diferencian tres variedades (Krippendorff, 1980, 1990) de *fiabilidad* en el *análisis de contenido*:

- a) *Estabilidad*: grado al que varían los resultados de la clasificación a lo largo del tiempo. Siempre que se emplee la misma agrupación de *categorías*, han

de alcanzarse los mismos resultados en la clasificación de las *unidades de codificación*. Por lo que se debe eliminar cualquier ambigüedad en la clasificación de las *unidades* en categorías.

- b) *Reproducibilidad*: grado al que se alcanzan los mismos resultados, cuando un mismo texto se codifica por varias personas. En este caso, la *infiabilidad* se atribuye a arbitrariedades en la *codificación*. Éstas pueden deberse a: ambigüedades en las instrucciones de codificación, errores de grabación o a diferencias cognitivas entre las personas que realizan la *codificación*.
- c) *Adecuación*: grado al que la clasificación del documento se corresponde con un estándar o norma. Este último criterio apenas se aplica porque rara vez se establecen *codificaciones* estándar.

La *fiabilidad* puede calibrarse mediante dos coeficientes principales (Sánchez Carrión, 1985):

- La *razón de codificaciones de acuerdo*, entre el total de codificaciones:

$$C.R. = \frac{2M}{N_1 + N_2}$$

Donde: “M” es el número de codificaciones en las que coinciden los 2 codificadores.

“N₁” y “N₂”, el número de codificaciones efectuadas.

- *Índice de fiabilidad* (p_i) de Scott:

$$P_i = \frac{\text{proporción de acuerdo observado} - \text{proporción de acuerdo esperado}}{1 - \text{proporción de acuerdo esperado}}$$

Su valor oscila entre 0,00 (desacuerdo total) y 1,00 (acuerdo total entre los codificadores).

EJEMPLO DE ANÁLISIS DE CONTENIDO

Análisis de contenido realizado por Juan José Igartua Perosanz, Carlos Muñiz Muriel, José Antonio Otero Parra y Montse de la Fuente Juan (2007: “El tratamiento informativo de la inmigración en los medios de comunicación españoles. Un análisis de contenido desde la Teoría del Framing”, *Estudios sobre el mensaje periodístico*, 13: 91-110).

Analizan el tratamiento informativo de la inmigración en la prensa y la televisión de ámbito nacional partiendo de tres *hipótesis* (a partir de estudios previos):

1. Un mayor número de noticias sobre acontecimientos negativos, en detrimento de las noticias sobre acontecimientos de carácter positivo.
2. En las noticias sobre inmigración predominarán los enfoques negativos de la inmigración, que destacan los *problemas* de la inmigración y la vinculan al aumento de la delincuencia.
3. Un mayor sensacionalismo en el medio televisivo frente a la prensa.

Muestra. La recogida de datos duró 6 meses (de enero a junio de 2004). Durante ese período de tiempo se seleccionó una *muestra* de los informativos televisivos emitidos en *prime time* en las cadenas TVE1, Antena 3 y Tele 5. Al mismo tiempo se seleccionó una *muestra* de diarios de información general de ámbito nacional (*El País*, *El Mundo*, *ABC*, *La Razón*). En cada mes se revisó una semana completa (eligiendo de forma aleatoria cada día de la semana). Se exploraron un total de 126 informativos televisivos (tres diarios por día) durante un período de 42 días (7 días cada mes) y 168 diarios (cuatro por día durante el mismo período de tiempo).

La primera fase de la investigación consistió en la *selección e identificación* de noticias sobre inmigración e inmigrantes dentro de cada informativo (que habían sido previamente grabados en vídeo) y periódico (en la hemeroteca de la universidad). Se identificaron y analizaron un total de 458 noticias (133 en televisión y 325 en prensa).

Libro de códigos. Se elaboró para determinar qué variables serían analizadas en cada una de las unidades de análisis. Este libro incluyó los siguientes apartados:

- *Importancia y emplazamiento de la información.* En el caso de las noticias de prensa, se evaluó: si la información aparecía en la portada del diario (1 = sí, 0 = no); si dicha información abría la sección (1 = sí, 0 = no); si aparecía en página par (0) o impar (1); el tamaño de la noticia (1 = menos de un cuarto de página, 2 = entre uno y dos cuartos de página, 3 = entre 2 y 3 cuartos de página, 4 = más de tres cuartos de página, pero no completa, y 5 = página completa); la presencia (1) o ausencia (0) de fotografías; la presencia (1) o ausencia (0) de otros elementos gráficos (como esquemas, gráficos o infografías).

Un análisis factorial de *componentes principales* en las seis variables extrajo dos dimensiones latentes (que explican el 58,4% de la varianza). El 1.º factor lo componen las variables: abrir la sección, presencia de otros elementos gráficos, tamaño de la noticia y aparición de la noticia en la portada. El 2.º factor: presencia fotográfica de la noticia. A partir de estos datos se creó una nueva variable, a la que se llamó “índice de importancia de la noticia dentro del diario”, formada por las cuatro variables que integraban el 1.º factor (con un coeficiente *alpha* de Cronbach de “0,74”, confirmando el carácter unidimensional del índice creado).

En el caso de las noticias emitidas en televisión, se computó la duración de la noticia (en segundos). Se evaluó si se recogía la noticia en el sumario o en titulares (1 = sí, 0 = no), la posición de la noticia dentro de la sección (1 = si aparecía en primer lugar, 0 = si aparecía en otros lugares) y la posición de la noticia dentro del informativo en su conjunto (1 = si era la primera noticia del informativo, 0 = si ocupaba otra posición u orden). Un análisis factorial de *componentes principales* con es-

tas cuatro variables concluye con un único factor (que explica el 51,91% de la varianza). A partir de ellas se crea el “índice de importancia o relevancia de la noticia”.

- *Enfoques temáticos de la inmigración:* se evaluaron los encuadres temáticos de cada noticia mediante una escala compuesta por 30 ítems que medían, de forma dicotómica, la presencia (1) o ausencia (0) de diferentes aspectos relacionados con el acontecimiento principal de la noticia. La escala se elaboró a partir de una investigación previa.
- *Carácter evaluativo del acontecimiento principal de la noticia:* se valoraba si el suceso o hecho narrado tenía carácter positivo (3), negativo (1), neutro o ambiguo (2), desde la perspectiva de los inmigrantes. Un ejemplo de noticias de carácter positivo (para el inmigrante) son las relacionadas con su regularización; de carácter negativo, la muerte o detención de inmigrantes; y neutro, si en el relato de los hechos no se destacaban consecuencias negativas ni positivas para los inmigrantes.
- *Análisis de las fotografías e imágenes de vídeo:* en el caso de la prensa, se analizó el tipo de imágenes de los inmigrantes que aparecían en las fotografías que acompañaban a las noticias. Con respecto a la televisión, las imágenes de los inmigrantes que aparecían en la pieza informativa. Se facilitó un listado con la descripción de 8 tipos de imágenes sobre inmigrantes: mujeres inmigrantes, inmigrantes heridos o muertos, inmigrantes dando testimonio o informando, inmigrantes niños o jóvenes, inmigrantes detenidos, inmigrantes trabajando, inmigrantes en situación de ocio o diversión, y representantes de asociaciones u organización de los inmigrantes. Los codificadores debían evaluar la presencia (1) o ausencia (0) de cada una de estas imágenes.

Codificación. Después de efectuar el proceso de codificación, se realizó un nuevo análisis con el 50% de las noticias seleccionadas aleatoriamente (67 noticias de televisión y 169 de prensa), con el objetivo de calcular la *fiabilidad* del proceso de codificación. Se utilizó la fórmula P_i de Scott (antes comentada), obteniendo un valor medio (0,73).

Resultados. Se observó que la importancia otorgada a las noticias sobre inmigración era mayor en el medio televisivo (*media* = 1,27) que en la prensa (*media* = 0,41); el 65,9% de las noticias se referían a acontecimientos de carácter negativo y el 23,6% positivo, habiendo diferencias estadísticamente significativas entre prensa y televisión (78,2% de las noticias de televisión frente al 60,9% de las de prensa se referían a acontecimientos de carácter negativo). Respecto a los encuadres noticiosos, el 43% de las noticias utilizaba un único encuadre (56,4% de las noticias de televisión, frente al 37,5% de las noticias de prensa). Tenía una presencia dominante el encuadre que vinculaba la inmigración con la delincuencia, las mafias y las acciones terroristas (*media* = 0,76; *desviación típica* = 1,09). Mediante los estadísticos *lambda de Wilks* y *F* se analizó la significatividad de las diferencias de los encuadres noticiosos en prensa y televisión.

En los análisis de las imágenes de las noticias de televisión se observó que la aparición de mujeres inmigrantes (33,3%), de inmigrantes heridos o muertos (25,8%), detenidos (20,5%) y de niños o jóvenes (19,7%) constituían los encuadres visuales más característicos de este medio. En las noticias de prensa predominaban las imágenes de mujeres (20,1%), de niños o jóvenes (9%), de inmigrantes detenidos (8,3%), heridos o muertos (8,3%) y, en menor medida, de inmigrantes trabajando (4,2%).

Para comprobar qué imágenes se utilizaban en las noticias que tenían una colocación privilegiada en los telediarios y en los diarios, se realizó un análisis de *correlación* entre

el índice de importancia de las noticias y las imágenes de inmigrantes reflejadas en las imágenes. Las noticias que recibían una colocación más privilegiada mostraban imágenes de inmigrantes heridos o muertos e imágenes de inmigrantes detenidos.

4.3. Evaluación de los datos secundarios

La investigación a partir de datos *secundarios* ha de cumplir similares criterios de *validez* y *fiabilidad* que el análisis de datos *primarios*. A lo dicho en el capítulo 2 y el subapartado anterior (referido al análisis *de contenido*), hay que sumar la especificidad que corresponde a los *datos secundarios*. A este respecto la evaluación de la *calidad* de los datos ha de cubrir –siguiendo a Stewart (1984)– los apartados siguientes:

- a) *Fuentes de datos*: ¿cuál era el propósito del estudio?, ¿quién recogió la información?
- b) *Metodología empleada* cuando se recopilaron los datos. Específicamente, se precisa información sobre:
 - *Diseño muestral*: representatividad de la *muestra*.
 - *Porcentajes de respuesta* conseguidos.
 - *Técnicas de obtención de datos* aplicadas (observación, entrevista, documentación).
 - *Técnicas de análisis* utilizadas en la información reunida.
- c) *Tiempo de recogida de los datos*: fecha de obtención de la información. Siempre hay un desfase temporal entre las fechas de recogida de los datos y la publicación de los resultados. Para su análisis e interpretación es imprescindible conocer la fecha en la que se obtuvieron y poderlos contextualizar en un momento sociopolítico y económico concreto.
- d) *Adecuación de los análisis y conclusiones*: comprobar la consistencia de la información con la proporcionada por otras fuentes. La confianza en los datos aumenta, cuando la misma información se recaba de fuentes distintas. Por lo que se recomienda que, en la medida de lo posible, se contraste la información con al menos dos fuentes diferentes. En caso de discrepancias, habría que identificar posibles razones y determinar la fuente que se estima más fiable.

De acuerdo con MacDonald y Tipton (1993) los *criterios de evaluación* pueden reducirse en cuatro esenciales:

- a) *Autenticidad*: si el documento es o no auténtico. En caso de serlo, si lo es en su totalidad o sólo en una parte.
- b) *Credibilidad* de los datos y de la fuente consultada.
- c) *Representatividad* respecto del conjunto de documentos existentes.

- d) *Significado* aparente y profundo. Para esto último se emplearían, preferentemente, técnicas analíticas cualitativas, como el *análisis estructural* o el *interpretacional*.

Tras la evaluación, se procede al *análisis* de la información. Como se ha visto en los apartados anteriores, el análisis puede quedar restringido a una mera *lectura analítica* (cuando se realizan *revisiones* de investigaciones) o dar un paso más allá, en la forma de *análisis de contenido* (cuantitativo o cualitativo) y *meta-análisis*, por ejemplo. En ambos casos se aplicarían técnicas de análisis estadístico, al igual que en el análisis de datos *primarios*: desde análisis univariantes *exploratorios* hasta *multivariantes*, que se resumen en el capítulo 8 (a cuya lectura se remite).

4.4. Ventajas e inconvenientes de la investigación a partir de datos disponibles

Como colofón al capítulo, el cuadro 4.2 sintetiza las ventajas e inconvenientes principales de la investigación con datos *secundarios*. Sin duda la principal *ventaja* está en su disponibilidad y bajo coste, especialmente si la información se halla publicada. En caso contrario, el acceso a dicha información puede resultar arduo. En general, el recurso a esta estrategia de investigación se recomienda cuando:

- a) Se desea información sobre acontecimientos del pasado, y ésta es difícil de obtener por otros medios (por no localización de sus actores, o porque éstos no la recuerden con exactitud).
- b) Se diseña una *investigación primaria*. La *revisión bibliográfica* es imprescindible para concretar el *problema* de investigación y su operacionalización: formulación de *hipótesis*, acotación de la *población* a analizar, selección de la *metodología* a emplear (a partir de la experiencia alcanzada en investigaciones precedentes).
- c) Se dispone de recursos (económicos, materiales y humanos) limitados para desarrollar una *investigación primaria* (y su correspondiente *trabajo de campo*).
- d) Se precisa dar una visión general de un fenómeno social concreto, incluyendo diferentes entornos socioculturales y períodos de tiempo.

La *información secundaria* es de gran utilidad en *estudios comparativos* (nacionales e internacionales) y de *tendencias* (ideal para el análisis *longitudinal*). Para ello se precisa que se hayan seguido los mismos criterios para la obtención de información en las fuentes originales (*por ejemplo*, aplicar el mismo *cuestionario* en los diferentes países o períodos de tiempo que se analizan). En caso contrario, no podrán compararse los resultados obtenidos.

- e) En el *diseño muestral*, al proporcionar los *datos censales* características de la *población* de interés. Esta información resulta necesaria para adecuar la *muestra* a la *población* y, de esta forma, alcanzar una mayor representatividad y precisión de las *estimaciones muestrales*.

- f) En los *diseños cuasi-experimentales*, cuando se comprueban los efectos de intervenciones o *tratamientos concretos*, antes y después de su aplicación.
- g) En cualquier *investigación multimétodo*, como complemento a la información recabada mediante otras *estrategias de investigación*.

En cuanto a sus *inconvenientes*, el principal es que la información “disponible” no se adecue a los objetivos de la investigación e incluso que no exista. Los *datos secundarios* pueden ser incompletos o haber sido recogidos para otros propósitos diferentes de los perseguidos por el investigador que los consulta. A la no adecuación se suman otros *inconvenientes* fundamentales:

- a) Como generalmente transcurre un largo período de tiempo desde que se recoge la información original y ésta se analiza y publica, suele haber una demora considerable en la disponibilidad de los *datos secundarios*. La demora puede incluso alcanzar los tres años, como sucede en las *estadísticas judiciales*, que elabora el INE. En las estadísticas de *Eurostat*, entre dos y tres años en función de la información aportada por los diferentes países de la Unión Europea. Esta demora restringe su adecuación para la descripción de fenómenos contemporáneos.
- b) Cuando se utilizan *datos secundarios* no siempre se tiene información que ayude a evaluar la calidad de los datos proporcionados. Concretamente, el conocimiento de los *errores muestrales*, de *medición* (operacionalización de conceptos, interpretación), y los de *obtención y análisis* de la información. Su conocimiento ayudará a evaluar la calidad de la información registrada y a su interpretación. De ahí la importancia de obtener una copia del instrumento de medición empleado, junto con las especificaciones técnicas de su aplicación.
- c) Las *fuentes secundarias* disponibles pueden reunir información únicamente de algunas de las variables que el investigador precise conocer. Además puede que estas variables hayan sido medidas de forma que no coincide con los intereses del investigador.

Los *datos estadísticos* suelen presentarse agrupados en un número concreto de categorías, y para una población determinada. Si el investigador deseara información individual o desagregada para un segmento específico de la población, por ejemplo, los *datos secundarios* no lograrían cubrir sus objetivos.

- d) Los *conceptos* también pueden cambiar de significado a lo largo del tiempo. Ello dificulta la comparación de datos reunidos en fechas diferentes. La comparabilidad también se agrava, cuando unos mismos *conceptos* no se han medido de la misma forma en distintas fuentes de *datos secundarios*.

Pese a estos inconvenientes, el *uso de datos secundarios* se muestra “imprescindible para buena parte de los planteamientos macrosociológicos, en los que se trata de indagar cuestiones referentes a la estructura social global o a la articulación de sus subestructuras” (Beltrán, 1989: 34-35). Por lo que algunos autores se muestran

bastante optimistas, y estiman que “el análisis secundario es probable que mantenga una posición dominante en la investigación de las ciencias sociales para el futuro previsible” (Kiecolt y Nathan, 1987: 11). Más si cabe en los últimos años, favorecido por la extensión de Internet y la digitalización de *archivos de datos*, muchos de ellos de acceso gratuito a través de la Red (caso del banco de datos del CIS y de la REIS, *por ejemplo*).

CUADRO 4.2. Ventajas e inconvenientes de la investigación con datos “secundarios”

VENTAJAS	INCONVENIENTES
<ul style="list-style-type: none"> – Precisa menos recursos económicos y temporales – Facilita el acceso a un mayor volumen de información – Permite cubrir amplios períodos de tiempo – Ayuda al diseño y realización de un estudio: desde los inicios de la investigación (objetivos, hipótesis...) hasta su conclusión (interpretación de los resultados) 	<ul style="list-style-type: none"> – No disponibilidad, bien debido a la dificultad o restricción en su acceso, o bien a su no adecuación a los objetivos de investigación, al período de tiempo o a la población analizada – Falta de control o de conocimiento del proceso de obtención y registro de la información. Ello limita la evaluación de la calidad de los datos y su interpretación – Su potencial de favorecer la comparación de los datos se restringe cuando acaecen cambios en la organización de la información a lo largo del tiempo

Lecturas complementarias

- Botella, J. y Gambará, H. (2002): *Qué es el meta-análisis*. Biblioteca Nueva. Madrid.
- Glass, G. V.; McGraw, B. y Smith, M. (1987): *Meta-analysis in social research*. Sage. Beverly Hills.
- González Rodríguez, B. (1994/2000): “La utilización de los datos disponibles”, en García Ferrando, M. *et al.*, *El análisis de la realidad social*. Alianza. Madrid: 245-273.
- Kiecolt, K. J. y Nathan, L. A. (1987): *Secondary analysis of survey data*. Sage. Beverly Hills.
- Krippendorff, K. (1990): *Metodología del análisis de contenido: teoría y práctica*. Paidós. Barcelona.
- Lee, R. M. (2000): *Unobtrusive methods in social research*. Open University Press. Londres.
- López-Aranguren, E. (1989/2000): “El análisis de contenido”, en García Ferrando, M.; Ibáñez, J. y Alvira Martín, F., *El análisis de la realidad social*. Alianza Universidad. Madrid: 383-414.
- Stewart, D. M. (1984): *Secondary research: information sources and methods*. Sage. Beverly Hills.
- Webb, E. J.; Campbell, D. T.; Schwartz, R. D.; Sechrest, L. y Grove, J. B. (1981): *Nonreactive Measures in the Social Sciences*. Houghton Mifflin. Dallas, TX.
- Weber, R. P. (1994): “Basic content analysis”, en Lewis-Beck, M. S., *Research practice*, Sage. California: 251-338.
- Weiss, S., Indurkha, N., Zhang, T. y Dameran, F. (2005): *Text mining: predictive methods for analyzing unstructured information*. Springer, Berlín.

Ejercicios propuestos

1. En un estudio sobre *discriminación*, ¿qué fuentes documentales y estadísticas consultaría? Realice una búsqueda bibliográfica e indique las palabras claves utilizadas.
2. Diseñe una investigación “no reactiva” sobre *xenofobia* en todas sus fases.
3. Describa una investigación en la que se haya aplicado análisis de *contenido* cuantitativo.
4. Se quiere analizar la influencia de la prensa escrita durante el desarrollo de una campaña electoral aplicando análisis de *contenido*. Diseñe la investigación, especificando la selección de la *muestra* de estudio, las unidades de codificación, de contexto y el plan de análisis.
5. Destaque las semejanzas y divergencias de los criterios de *validez* y *fiabilidad* aplicados en la investigación a partir de fuentes secundarias con los descritos en los capítulos anteriores.

5

LA ENCUESTA COMO ESTRATEGIA DE INVESTIGACIÓN SOCIAL

Hasta el presente la *encuesta* continúa acaparando el protagonismo en la metodología cuantitativa como estrategia de investigación social. A ello contribuyen sus amplias posibilidades para obtener información de un conjunto amplio de la población y en un laxo temporal relativamente breve. Más en los últimos años con la extensión de los modos de encuesta digitales, que han contribuido a su abaratamiento y a reducir el tiempo de recogida y de análisis de los datos recabados. Pero su *calidad* y relevancia se halla supeditada al rigor de su diseño y ejecución. Como se irá desglosando a lo largo de éste y el capítulo siguiente, son varios los *errores* que pueden confluir en la realización de la encuesta, al igual que las actuaciones para su resolución. Pero comencemos por su fundamentación y caracterización metodológica.

5.1. Fundamentos de la encuesta

Aunque los orígenes de la encuesta comúnmente se sitúan en los primeros recuentos censales realizados con fines tributarios y militares (*Domesday Book* fue un famoso censo de Inglaterra realizado en 1085 por William *el Conquistador*), no es hasta el movimiento de los *reformadores sociales* cuando comienza las “encuestas sociales”. Un conjunto de profesionales (médicos, ingenieros, naturistas) unidos por un mismo propósito: aportar evidencia empírica de los problemas sociales que siguen al desarrollo de la Revolución Industrial y al rápido crecimiento de las ciudades. Su finalidad era eminentemente política: movilizar a la opinión pública para exigir medidas concretas de *reforma social*. De ahí su nombre: “reformadores sociales”.

La recogida de información a través de parroquias e inspecciones fiscales, que hasta entonces había imperado, se muestra insuficiente. Diferentes *sociedades estadísticas* que fueron creándose en varios países europeos (Inglaterra, Francia, Ale-

mania) promovieron la realización de “*encuestas sociales*”. Suponían la cumplimentación de un cuestionario *estandarizado*, más sobre aspectos objetivos (o de hechos) que subjetivos (actitudes, opiniones, valores). Incluían información variada, principalmente sobre las condiciones de vida y de vivienda de la clase trabajadora. También, su práctica religiosa, escolarización, hábitos de higiene y de salud, en general. Los entrevistadores, o personas responsables de su administración, obtenían la información tanto mediante la observación *directa* de los barrios obreros, como la proporcionada de manera *verbal* por los propios afectados o por terceros (empresarios, personal escolar, sanitario, religiosos). Entre estos estudios destacan:

- *Statistical Account of Scotland* (1791-1825): una monumental investigación, de 21 volúmenes, realizada por John Sinclair sobre los problemas de la población rural en Escocia. La información se obtuvo de un cuestionario diseñado ex profeso. Contenía 116 preguntas referentes a diversas características socio-demográficas de la población (sexo, edad, ocupación, mortalidad, práctica religiosa), de su entorno (orografía, recursos minerales y agrícolas) y de la actividad económica de la región (desempleo, producción agrícola, ganadera). Para su realización, el autor contó con la colaboración del clero escocés, que le proporcionó información de 881 parroquias de Escocia. Tuvo gran repercusión en el desarrollo de los censos de población (el primero data de 1801 en Inglaterra y Francia; en España, en 1857) y en la elaboración de cuestionarios (Kovaliova, 1989; Garrigós, 2003).
- *Labour and the Life of the People of London* (Macmillan and Co. Londres, 1902-1903), una investigación de 17 volúmenes financiada y realizada por Charles Booth (un acaudalado naviero británico), desde 1889 hasta 1897. Easthope (1974) la sitúa en los orígenes de la encuesta moderna. Esta investigación describe las condiciones de vida de la población trabajadora londinense. La información se obtuvo del uso combinado de distintas fuentes de datos: el *Censo de Población* británico de 1881; *informes* policiales, de organizaciones de beneficencia, de inspectores escolares y sanitarios; la *observación directa*, por el mismo autor, de varios barrios obreros, donde vivió durante algún tiempo; la realización de *entrevistas* a hombres de negocios y a asistentes escolares, para indagar en las condiciones de vida de los escolares y de sus familias. Además, procedió a la selección aleatoria de representantes de familias de los diferentes grupos sociales, excluyendo al inferior.

A Booth se le reconoce el acierto de perfeccionar “la metodología de sondeos, rechazando la ayuda de los intermediarios –observadores de la clase media– y apelando directamente a los obreros, cuya situación estudiaba” (Kovaliova, 1989: 131). El hecho de haber entrevistado directamente a las familias le convierte en “uno de los primeros en llevar a cabo una encuesta mediante entrevista cara a cara” (De Leeuw y Collins, 1997: 199).

- *Poverty. A Study of Town Life* (Macmillan, 1906), una investigación similar a la anterior, que realiza uno de los seguidores de Booth: Benjamin Seebohm

Rowntree. El cuestionario igualmente incluyó información descriptiva referida a las condiciones de trabajo (salarios, jornada laboral) y de vida (características de las viviendas, condiciones de higiene, de salubridad, necesidades alimenticias), de las clases sociales más desfavorecidas en la Inglaterra de finales del siglo XIX. Distinguió la pobreza primaria (familias que carecían de recursos para cubrir las necesidades mínimas) de la secundaria (sí disponían de recursos suficientes, pero los malgastaban, quedándose en la pobreza).

- *Les Ouvriers Européens* (su primera edición, de 36 monografías, data de 1855; la segunda, en seis volúmenes, de 1877 a 1879, e incluye 57 monografías de familias), de Frédéric Le Play. Una investigación calificada por Nisbet (1966: 61) como “el primer trabajo sociológico genuinamente científico del s. XIX”. Destaca por la elección de la familia como unidad de observación. Describe las características externas y “morales” (sentimientos, valores, deseos) de varias familias obreras (de trabajadores de la industria, la agricultura, la artesanía y el comercio), con las que convivió el autor y sus colaboradores (también ingenieros de minas), en sus viajes por varios países europeos (España, Inglaterra, Francia, Suiza, Austria, Hungría, Prusia, los Países Nórdicos e, inclusive, Rusia). Un total de 22 familias, en cuya selección se primó que fuesen “representativas” de una población concreta. Los “encuestadores” (o colaboradores) debían permanecer con la familia de 8 a 10 días. Durante ese tiempo de convivencia cumplimentaban un cuestionario con preguntas referidas a la vivienda (muebles, utensilios, animales domésticos, propiedades familiares), los hábitos alimenticios y de trabajo de los miembros de la familia. Esta información recabada mediante *observación directa* se complementó con la obtenida de la interacción y *entrevista* a los integrantes de la familia. Además se interrogó a personalidades de la localidad, que fueron de gran ayuda para la selección de las familias “representativas”.

Karl Marx fue otro de los defensores de extraer la información de los propios trabajadores, para conocer sus condiciones de vida y de trabajo. A él se le atribuye la primera *encuesta por correo*, que data de 1880. Cuando envía 25.000 cuestionarios a obreros franceses. Pensaba que éstos eran quienes mejor podían describir sus condiciones sociolaborales. No obstante, la respuesta a esta primera iniciativa fue mínima. La investigación quedó paralizada.

Más éxito obtuvo Max Weber, otro de los padres fundadores de la sociología. Su obra fue determinante para el desarrollo de la metodología cualitativa, pero también realizó una importante contribución empírica al desarrollo de la *encuesta*. Principalmente, en dos aspectos metodológicos claves: la correcta *selección de los informantes* y la mejora del *cuestionario*. Lazarsfeld y Oberschall (1965) destacaron cuatro aportaciones principales:

1. Su participación en una investigación sobre las condiciones del trabajo rural en Alemania, promovida por la *Verein für Sozialpolitik*, una asociación crea-

da en 1892 por un grupo de profesores universitarios (preocupados por el creciente antagonismo de los trabajadores alemanes hacia el Estado alemán). Además de solicitar reformas sociales, querían minimizar la influencia del pensamiento marxista en los trabajadores. Algo más de 3.000 terratenientes recibieron un cuestionario detallado para que describiesen su situación particular. El 70% lo respondió. Asimismo, se enviaron 600 cuestionarios a personas que se creía ofrecerían una visión más general. Aproximadamente el 50% lo devolvió relleno. Weber realizó el informe de Prusia del Este, en 1892, y fue el único investigador que hizo un *análisis comparativo*. Comparó sus datos con los de estudios anteriores para darles perspectiva histórica. De este estudio Weber criticó que sólo se utilizase, como informantes, a los terratenientes y el contenido del cuestionario aplicado: demasiado énfasis en las condiciones materiales de los trabajadores y escaso análisis de los aspectos subjetivos.

2. En 1893 Weber lleva a cabo otra encuesta, sobre trabajadores del campo, encargada por el Congreso Social Evangélico. Éste demandaba un mayor conocimiento de las condiciones de vida de los grupos de población de renta más baja. Envío 10.000 cuestionarios a clérigos por dos razones principales: una, el Congreso tenía un registro central de todas las parroquias; dos, pensaba que éstos podrían mejor informar de problemas psicológicos. El cuestionario fue más breve que el de la encuesta anterior. Se centró más en la movilidad social y laboral, y en los orígenes de los distintos grupos de campesinos. También quiso conocer las actitudes de los trabajadores hacia formas no tradicionales de trabajo. Le fueron devueltos 1.000 cuestionarios. Los resultados de esta investigación se publicaron en periódicos y en informes del Congreso Social Evangélico.
3. En 1907 Weber participó en una serie de investigaciones, efectuadas por *Verrein für Sozialpolitik*, sobre la selección y adaptación de los trabajadores en las grandes industrias. Esta vez la información se obtuvo combinando *datos disponibles*, en las oficinas de las fábricas seleccionadas, con *observación participante* y *entrevistas* “directas” a los trabajadores. Para estas entrevistas se diseñó un cuestionario de 27 preguntas relativas a la ocupación de los padres, actividades de tiempo libre, elección ocupacional o metas en la vida.
4. Sobre esta misma temática, en 1908 Weber realiza un estudio detallado de la productividad de los trabajadores en una fábrica textil, que pertenecía a la familia de su esposa. Los resultados los publica en un informe que titula *La consideración de los aspectos psicofísicos del trabajo en la industria*. Este estudio destaca por su cuidadoso análisis estadístico de los resultados: la tabulación cruzada. También, porque demuestra que las personas de renta baja pueden ser “buenos” informantes.

El estudio de Adolf Levenstein de 1912, *La cuestión obrera*, igualmente ilustra el uso de la encuesta a la población trabajadora en la investigación empírica alemana. Describe la situación de los obreros de la industria alemana, a partir de los re-

sultados de una *encuesta por correo* enviada a 8.000 obreros, de tres categorías (de la industria minera, metalúrgica y textil), de ocho regiones industriales, a razón de 1.000 cuestionarios por cada región, desde 1907 a 1911. Obtuvo un porcentaje de cuestionarios devueltos elevado (el 63% de los enviados) para aquel entonces. No obstante, la encuesta carecía de rigor desde la óptica actual. Primero, los cuestionarios se enviaban a amigos y conocidos entre los obreros. Después, a quienes respondían, se les pedía que los distribuyesen entre sus conocidos o amigos. En su análisis, Levenstein siguió la recomendación de Weber, publicada en su artículo, de 1909, “Zur Methodik Sozialpsychologischer Enqueten und ihrer Bearbeitung” (en castellano: “Sobre la metodología de las encuestas sociopsicológicas y sobre su análisis”) de realizar un análisis estadístico de sus datos en la búsqueda de *tipologías* para clasificar a los obreros alemanes. Lazarsfeld y Oberschall (1965) lo sitúan como un trabajo pionero en el estudio de las actitudes, aunque ni Levenstein ni Weber hicieron explícita la idea de que las actitudes pueden medirse y que las variables presentes en los cuestionarios pueden combinarse para clasificar a las personas.

5.1.1. Configuración de la encuesta “por muestreo”

Ninguna de las “*encuestas sociales*” expuestas se adecuan a lo que actualmente se entiende por *encuesta*: “*encuesta por muestreo*”. Más bien se asemejaban a *censos*, que pretendían cubrir a las poblaciones totales de las comunidades que analizaban. Aunque contribuyeron al conocimiento empírico de la sociedad de su tiempo (la población trabajadora, en concreto), el avance de la *encuesta* precisó de aportaciones de diferentes ámbitos de conocimiento.

En 1883, Galton delimita (en *Inquiries into the Human Faculty*) algunos requisitos básicos de los *cuestionarios*. Éstos han de cubrir los objetivos de la investigación siendo, al mismo tiempo, fáciles de entender y de contestar. Pero, los beneficios más notorios proceden del área de la estadística, de la aplicación de la *teoría de la probabilidad* en la selección de las unidades de observación:

- Arthur L. Bowley, el primer profesor de estadística en la *London School of Economics and Political Science*, fue quien introdujo el *muestreo probabilístico* en la práctica de encuesta. A él se le atribuye la primera encuesta que se realiza mediante *selección aleatoria* de los informantes. Una investigación que realiza junto con A. R. Burnett-Hurt, publicada en 1915 con el título *Livelihood and Poverty* (en la editorial londinense Bell). En publicaciones precedentes había defendido la necesidad de aplicar *muestras aleatorias representativas* en la investigación mediante encuesta. Destacan sus artículos: “Working class households in Readings”, de 1913 (*Journal of the Royal Statistical Society*, 76: 672-691), y “Presidential address to the Economic Section of the British Association for the Advancement of Science”, de 1906 (*Journal of the Royal Statistical Society*, 69: 540-558). También su libro sobre la medición

de los fenómenos sociales, de 1915, *The Nature and Purpose of the Measurement of Social Phenomena* (P. S. King and Son, Ltd., Londres). Contribuyó a la definición de los parámetros de calidad de los datos de encuesta, además del desarrollo de la *encuesta por muestreo*.

- A. N. Kiaer (el director de *Norwegian Bureau of Statistics* de Oslo) fue quien primero defendió la exigencia de que la encuesta se hiciera sobre *muestras representativas*, en su libro *Representative Method of Statistical Surveys*, publicado en 1897 (en la editorial Kristinia, de Oslo). En él expone que, de la información extraída de una *muestra* relativamente pequeña de la población, pero cuidadosamente seleccionada mediante algún procedimiento aleatorio, pueden estimarse los parámetros poblacionales, sin necesidad de elaborar un censo completo de la población, como hasta entonces se creía. Para ello la *muestra* ha de ser “representativa” (una representación a pequeña escala de la población de estudio). Asimismo propone el procedimiento de la *replicación* para evaluar los resultados de encuesta: extraer series de submuestras comparables repitiendo la operación de muestreo. En 1895 ya introdujo el término *investigación representativa* en un informe para el Instituto Nacional de Estadística (ISI: *International Statistical Institute*) que, a finales del siglo XIX, era el centro de debate para estadísticos oficiales.
- Jerzy Neyman desarrolla la propuesta de Kiaer de aplicar la *estratificación* en la selección aleatoria de la *muestra* en un artículo publicado en 1934: “On the different aspects of the representative method: the method of stratified sampling and the method of purposive selection” (*Journal of the Royal Statistical Society*, 97: 558-606). En él desarrolla la *estratificación óptima*, el muestreo por *conglomerados*, el muestreo en *poblaciones finitas*, la *distribución muestral* y el *error de muestreo*. Demuestra que el *error de muestreo* puede medirse calculando la varianza del estimador.
- La *replicación* adquiere un nuevo desarrollo de la mano del estadístico hindú P. C. Mahalanobis. En dos de sus publicaciones principales, de 1944 (“On the large scale sample surveys”. *Royal Society Phil. Trans. B.*, 231: 329-451) y de 1946 (“Recent experiments in statistical sampling in the Indian Statistical Institute”. *Journal of the Royal Statistical Society*, 19: 325-378), aboga por el uso de la *replicación* (a lo que denomina “interpenetrar muestras”), para medir la precisión de estimaciones derivadas de una *encuesta*. Además enfatiza la necesidad de incluir los “errores humanos” (la variabilidad introducida por entrevistadores, codificadores y supervisores), junto a los de muestreo, en la estimación de la precisión de una encuesta.

La aportación de estos estadísticos es continuada por la publicación de célebres manuales de muestreo (descritos en el apartado 7.1), que avanzan en la medición de los efectos del diseño *muestral* en la imprecisión de las estimaciones de encuesta.

Desde la psicología, la aportación al desarrollo de la investigación mediante encuesta se focaliza en la *medición* de los *conceptos* teóricos, en su *fiabilidad* y *validez*.

De ello, al igual que de la medición de las actitudes, se ha hablado en el capítulo 2. A lo dicho se añade la relevancia que, para el diseño de los cuestionarios, ha tenido la aplicación de experimentos sobre los efectos debidos a la formulación de preguntas (enunciado, opciones de respuesta, ubicación en el cuestionario). Su relevancia puede verse en revistas especializadas, como *Public Opinion Quarterly* (creada en 1947, bajo la promoción de la organización profesional *The American Association for Public Opinion Research*), donde dominan experimentos dirigidos a la mejora de la calidad de la encuesta. Si bien, entre los pioneros cabe destacar los realizados por:

- A. Hobson (1916): “The use of the correspondence method in original research”. *Journal of the American Statistical Association*, 15: 210-218. Contrasta diferentes tipos de cuestionarios y su formulación en encuestas por correo.
- B. Muscio (1917): “The influence of the form of a question”. *The British Journal of Psychology*, 8: 351-389.

5.1.2. Dominio de la encuesta por muestreo en la investigación social norteamericana

A lo largo de los años veinte y principios de los treinta (del siglo XX) comienzan a aparecer, en Estados Unidos, empresas de investigación de mercados para la investigación comercial y la realización de sondeos preelectorales. Precisamente fue el acierto en la predicción de un resultado electoral, las elecciones presidenciales de 1936, lo que supuso un fuerte impulso a la *encuesta por muestreo*. Una macroencuesta, por correo, para la revista *The Literary Digest* (que defendía la práctica de a mayor número de personas encuestadas, mejor es la predicción), a miles de suscriptores con teléfono, fracasó en su predicción electoral, frente a una encuesta más modesta bajo la dirección de Gallup y Crossley. Esta última encuesta predijo la victoria de Franklin D. Roosevelt tras sólo haber entrevistado a 1.500 votantes norteamericanos, elegidos al azar, y establecidas unas *cuotas* por edad y sexo. Mostró que una *muestra* pequeña, pero cuidadosamente extraída, podía reflejar la realidad social mejor que una gran encuesta de “una muestra parcial, con poco o ningún esfuerzo para alcanzar porcentajes de respuesta razonables” (Rossi, Wright y Anderson, 1983: 5).

Los encuestados para *The Literary Digest* no fueron aleatoriamente seleccionados, del conjunto de electores norteamericanos, sino extraídos del listín telefónico de suscriptores a la revista y de propietarios de automóviles. Ello introdujo un importante *sesgo* en los resultados del sondeo. Según estimaciones de Biemer y Lyberg (2003), en aquellas fechas sólo un 35% de los hogares norteamericanos disponían de teléfono. Además, esta población se caracterizaba por ser desproporcionadamente republicana. Lo que explica la preferencia de sus sondeados por el republicano Alf Landon. Se excluyó a las personas de nivel de renta más bajo, que fueron quienes más votaron por el programa de recuperación económica defendido por Roosevelt.

A este importante *error de cobertura* hay que sumar otros, igualmente decisivos, como el relacionado con el bajo porcentaje de respuesta obtenido. Tan sólo un 22% de los diez millones de cuestionarios enviados fueron contestados y devueltos por correo.

A partir de entonces, los sondeos realizados por Gallup, Crossley y Ropper adquirieron un mayor protagonismo en el análisis de la opinión pública. Comenzaron a expandirse los institutos de opinión pública en Estados Unidos y Gran Bretaña. Gallup volvió a acertar en su pronóstico electoral (con *muestreo por cuotas*) en las elecciones presidenciales norteamericanas de 1940 y 1944, pero falló en 1948. Principalmente se debió a la no correspondencia de los datos de población registrados en el *Censo de Población* de EE UU de 1940 (utilizado para la extracción del *muestreo por cuotas*), con la población que realmente residía en el país en 1948. La Segunda Guerra Mundial afectó a la pérdida de población y al desplazamiento de habitantes de los municipios rurales a los urbanos. Por lo que los habitantes de municipios más rurales quedaron sobrerrepresentados en la *muestra* frente a los urbanos; y fue en las ciudades donde más se votó por el candidato demócrata.

Entre 1936 y 1949 Gallup realizó en torno a 400 *experimentos* sobre los efectos en la respuesta debidos a la redacción de la pregunta en el cuestionario. Muchos de estos experimentos fueron diseñados por Hadley Cantril, quien en 1940 crea *The Office of Public Opinion Research*, en la Universidad de Princeton. Quería que éste se convirtiera en centro de referencia para el estudio de la opinión pública y las técnicas de análisis, además de archivo de sondeos de opinión a disposición de cualquier investigador interesado. En 1944 publica (en la editorial de la Universidad de Princeton), junto a sus colaboradores, un compendio de estudios sobre el diseño de encuestas, su ejecución y análisis, con el título *Gauging Public Opinion*. Destaca su tratamiento sobre la formulación de las preguntas, su significado, el efecto de las baterías de preguntas, la medición de la intensidad de las opiniones, y los *sesgos* debidos a las características y la formación de los entrevistadores.

En las mismas fechas Paul F. Lazarsfeld funda *Bureau of Applied Social Research*, en la Universidad de Columbia; con anterioridad había creado *The Office of Radio Research*, en la Universidad de Newark. Y en 1944, Lazarsfeld publica (en la editorial de la Universidad de Columbia) una investigación que había realizado junto con H. Gaudet y B. Berelson: *The People's Choice: How the Voter Makes up his Mind in a Presidential Campaign*. Esta publicación adquiere una gran trascendencia en la investigación social por ser pionera en tres aspectos claves para la investigación mediante *encuesta*:

- a) El estudio del voto mediante *encuesta por muestreo* (desde la sociología académica).
- b) La aplicación de una *encuesta panel*: un mismo cuestionario se pasa a la misma muestra de individuos en diferentes fechas. El estudio tenía por objetivo principal analizar las intenciones de voto de los electores del condado de Erie (Ohio), en las elecciones presidenciales de 1940. Para ello, los autores

pasaron tres veces un mismo cuestionario a una *muestra* de electores extraída al azar. La primera vez, antes del nombramiento de los electores; la última, momentos antes de la elección. Mediante este seguimiento de la *muestra* querían comprobar el efecto de la campaña electoral en la intención de voto: las mismas personas son seguidas en el transcurso de la campaña electoral. A su vez, utilizaron otras tres muestras como grupos de *control*, con el propósito de comprobar los efectos de aplicar varias veces el mismo cuestionario a la muestra principal.

- c) El *análisis de datos* de encuesta: la introducción de las *tablas cruzadas* (de más de dos variables) para la confección de *tipologías*. Si bien, Rossi, Wright y Anderson (1983) critican que no aplicaran modelos de inferencia estadística (ninguna prueba de *significatividad estadística* ni de estimación de *intervalos de confianza*) y el procedimiento de muestreo seguido: “elegir una de cada cuatro viviendas”.

En EE UU, también durante la Segunda Guerra Mundial, se crea *The National Opinion Research Center* (NORC) y un departamento de investigación del ejército norteamericano, que estuvo bajo la dirección de Samuel A. Stouffer desde 1941 hasta 1945. Entre sus colaboradores destacaron los sociólogos Louis Guttman, Edwards A. Suchman, Robin Williams, Shirley Star o John Clausen, y los psicólogos Carl Hovland, Arthur Lumsdaine y William Bennet, quienes después proseguirían sus investigaciones en diferentes departamentos universitarios. Entre las investigaciones realizadas por este equipo investigador destaca, por su gran envergadura, un estudio sobre la moral, las opiniones y las actitudes del soldado americano durante la Segunda Guerra Mundial: *The American Soldier* (1949). Sus resultados fueron publicados (por la editorial de la Universidad de Princeton) en cuatro volúmenes (de 1947 a 1950), bajo el título genérico de *Studies in Social Psychology in World War II*. Aunque su propósito inicial fue eminentemente práctico (proporcionar información a las autoridades gubernamentales sobre las opiniones y actitudes de su personal militar), acabó teniendo una gran repercusión en el desarrollo de la metodología de encuesta. Principalmente destacó:

- a) Su contribución al diseño de las *encuestas autocumplimentadas* (a una *muestra* aleatoria de 5.000 soldados se les pidió que rellenasen un cuestionario diseñado al efecto).
- b) Formular varias preguntas para medir un mismo concepto teórico.
- c) El análisis de los datos de encuesta.

Robert K. Merton y Paul F. Lazarsfeld (1950) se refieren a esta investigación como el primer trabajo que enseña al lector no especializado (en la investigación sociopsicológica) a aprender los detalles de los métodos más avanzados en el campo, relativamente nuevo, de las encuestas de actitudes. Platt (1996) destaca su contribución a la consideración de la encuesta como el método líder en la recogida de información,

pese a los contenidos dados por Hovland, Lumsdaine y Sheffield, y a los comentarios repetidos de Stouffer sobre la superioridad del método *experimental*. En ello tuvo mucho que ver la monografía de Merton y Lazarsfeld (1950) sobre dicha investigación, por la orientación que se hizo de la *encuesta* y de su accesibilidad.

En los años cincuenta aparecen dos monografías sobre la *entrevista*, que consolidan los fundamentos de la entrevista de encuesta (llamada *entrevista estandarizada*):

- H. Hyman y colaboradores (1954: *Interviewing in Social Research*. Chicago University Press). Analiza cómo las expectativas del entrevistador afectan a los datos de encuesta.
- R. L. Kahn y C. F. Cannell (1957: *The Dynamics of Interviewing: Theory, Techniques and Cases*. John Wiley & Sons. Nueva York). Destaca la responsabilidad del *entrevistador* en el resultado de la encuesta. De él se espera que sea un profesional, que se limite a leer las preguntas del cuestionario como fueron redactadas, sin implicar ningún favoritismo, sorpresa o satisfacción hacia una respuesta concreta.

Pero no fueron éstas las primeras monografías sobre la *entrevista de encuesta*. En 1934 aparece una edición revisada (en la editorial Harper, de Nueva York) de *How to Interview*, de W. V. Bingham y B. V. Moore. En esta publicación ya se defendió la necesidad de *estandarizar* la actuación del entrevistador en la encuesta. También en artículos que analizan los efectos de la *actuación del entrevistador* en los resultados de encuesta, y que establecen pautas para su mejora. Entre los más referenciados están:

- S. A. Rice (1929: “Contagious bias in the interview”. *American Journal of Sociology*, 35: 420-423). Analiza posibles influencias de los entrevistadores que “contaminan” la encuesta.
- J. Wechsler (1940: “Interviews and interviewers”. *Public Opinion Quarterly*, 4: 258-260). Sobre el papel de los entrevistadores durante la entrevista.
- D. Katz (1942: “Do interviewers bias poll results?” *Public Opinion Quarterly*, 6: 248-268). Comprueba el efecto diferencial del estatus de los entrevistadores. Los llamados de “cuello blanco” registran una mayor proporción de actitudes y opiniones conservadoras que los de “cuello azul” o de clase trabajadora.
- D. Williams (1942: “Basic instructions for interviewers”. *Public Opinion Quarterly*, 6: 634-641). Detalla las instrucciones seguidas por los entrevistadores que trabajan para *The National Opinion Research Center* (NORC).
- M. T. Orne (1969: “Demand characteristics and the concept of quasi-controls”, en R. Rosenthal y R. L. Rosnow, *Artifacts in Behavioral Research*. Academic Press. Nueva York). Analiza cómo las características de la demanda afectan a la conducta de los entrevistadores y los entrevistados.

Ésta es la época de mayor expansión de la *encuesta* como estrategia metodológica. Coincide con los años de hegemonía del *funcionalismo* en la sociología norteamericana. Wells y Picou (1981) ponen cifras a dicha expansión: el 48,2% de los artículos publicados en la revista *The American Sociological Review*, de 1936 a 1949, utilizan la encuesta. Entre 1950 y 1964 el porcentaje asciende al 70,5%. Entre 1965 y 1978, hasta el 80,3%. Los métodos de *interpretación* descienden, por el contrario, del 50,4% al 27% y 17,1%, respectivamente. En la revista *American Political Science Review*, Wahlke (1979) contabiliza que la mitad de los 180 artículos sobre comportamiento político se basan en datos de encuesta. Un 20% complementa la encuesta con otras estrategias de investigación.

Es la culminación de un arduo proceso de desarrollo de la encuesta que se benefició de:

1. Los avances estadísticos (e informáticos) en los ámbitos del muestreo y del análisis de datos.
2. El desarrollo de escalas y de índices para medir (de manera sistemática) aspectos objetivos y subjetivos (opiniones, actitudes, valores) de la realidad social.
3. El surgimiento de la investigación de mercados.
4. El eco mediático de los datos de encuesta.
5. La competencia entre investigadores y universidades en busca de fondos privados y prestigio.
6. La creación de archivos o bancos de datos que permiten acceder a una amplia variedad de encuestas. Muchas universidades norteamericanas cuentan con centros de investigación sobre encuestas. Destacan: *National Opinion Research Center* (NORC), en la Universidad de Chicago; *Survey Research Center*, en la Universidad de California en Berkeley; *Institute for Social Research* (ISR), en la Universidad de Michigan.

Además se acrecienta la preocupación por mejorar la *calidad* de la encuesta, como puede apreciarse de las siguientes publicaciones de gran repercusión en el tratamiento de los *errores* de encuesta:

- J. L. Gillin (1915: “The social survey and its further development”. *Journal of the American Statistical Association*, 14: 603-610). Alerta de los peligros de la falta de control de la calidad de la encuesta.
- W. E. Deming (1944: “On errors in surveys”. *American Sociological Review*, 9: 359-369). Lista trece factores que pueden afectar a la calidad de la encuesta. Incluye los *errores* de muestreo, los debidos a la actuación del entrevistador, al procedimiento seguido en la recogida de datos, al diseño del cuestionario, a la *no respuesta*, a la codificación y procesamiento de la información, y los errores de interpretación.
- H. J. Parry y H. M. Crossley (1950: “Validity of responses to survey questions”. *Public Opinion Quarterly*, 14: 61-80). Sobre la *validez* de la medición mediante encuesta.

- M. H. Hansen, W. N. Hurwitz y W. G. Madow (1953: *Sample Survey Methods*, vol II. Wiley. Nueva York). Diferencia los *errores variables* de los *sesgos sistemáticos*, y propone el concepto de *error total de encuesta*.
- C. F. Cannell y F. J. Fowler (1963: “Comparison of a self enumerative procedure and a personal interview: a validity study”. *Public Opinion Quarterly*, 27: 250-264). Comprueba la *validez* de la entrevista personal frente al procedimiento “autoenumerativo”.
- S. Sudman y N. Bradburn (1974: *Response effects in surveys: a review and synthesis*. Aldine Publishing Co. Chicago). Un texto clave sobre los *efectos de respuesta*, en el que se proponen pautas o actuaciones para su resolución.
- H. Schuman y S. Presser (1981: *Questions and Answers in Attitudes Surveys*. Academic Press. Nueva York). A partir de investigación propia (experimentos), complementada por la llevada a cabo por otros investigadores, analizan los efectos debidos a la *forma de la pregunta*, su redacción y *contexto* en los resultados de encuesta.
- R. M. Groves (1989: *Survey Errors and Survey Costs*. John Wiley. Nueva York). Ofrece una revisión comprehensiva de los *errores* de encuesta y sus causas, relacionándolos con los costes de la investigación. Reconoce que el “coste” es un componente esencial de la calidad de los datos.
- L. E. Lyberg y colaboradores (1998: *Survey Measurement and Process Control*. John Wiley & Sons. Nueva York). Una compilación de artículos sobre las diversas fuentes de *error de medición* y sus efectos en la calidad de la encuesta.
- P. P. Biemer y L. E. Lyberg (2003: *Introduction to Survey Quality*. John Wiley & Sons. New Jersey). Sintetiza las diferentes fuentes de *error* de encuesta, insistiendo en su visión conjunta.

En los últimos años la literatura sobre metodología de encuesta ha crecido enormemente. Una muestra es la bibliografía del libro (en su mayoría limitada al mundo anglosajón e hispano), si bien es difícil recoger todo lo publicado hasta la fecha.

5.1.3. *Desarrollo de la investigación mediante encuesta en España*

En España los primeros antecedentes de la encuesta se hallan en la gran encuesta de ámbito nacional, promovida por la Sección de Ciencias Morales y Políticas del *Ateneo* de Madrid, y realizada entre 1901 y 1902. Trata de las costumbres populares y de los tres hechos más característicos de la vida: nacimiento, matrimonio y defunción. R. Salina fue el promotor de la encuesta. Contó con la colaboración de Puyol, Bernardo de Quirós, García Herreros, Pedregal y Camarón. Fueron devueltos al *Ateneo* 289 cuestionarios debidamente cumplimentados, distribuidos entre las distintas provincias españolas.

Si bien ya desde 1882, diferentes sociedades antropológicas españolas habían editado cuestionarios, de ámbito local, para ayudar a la recogida de datos (Lisón, 1968).

En 1895, J. Costa dirigió en el Ateneo *La encuesta sobre tutela social* y, en 1901, promovió (también en el Ateneo) *La encuesta sobre oligarquía y caciquismo como forma actual de gobierno en España*. Esta última encuesta contó con la colaboración de Pardo Bazán, G. Azcárate, Ramón y Cajal, Unamuno, Pí y Margall. En 1904 se crea el *Instituto de Reformas Sociales*. Desde él se promueven varios estudios “reformistas” sobre la sociedad española. Destaca la investigación relativa al trabajo en las minas, de 1910, o la primera encuesta a escolares, de Eugenio d’Ors, de 1921.

En la década de los cincuenta se produce un segundo desarrollo que se beneficia de la recepción de los avances de la metodología de encuesta (principalmente en EE UU). Entre las encuestas pioneras están:

- *La encuesta a estudiantes universitarios* de Madrid, realizada por Manuel Fraga Iribarne y Tena Artigas en 1949. Sus resultados se publican en 1953, en la *Revista de Educación* y la *Revista Internacional de Sociología*.
- *La encuesta a estudiantes universitarios* españoles, de José Luis Pinillos, de 1955. La primera que incluye preguntas de índole política.
- *La encuesta a estudiantes* de la Universidad de Valencia, de Francisco Murillo Ferrol y José Jiménez Blanco, de 1956.
- *La encuesta sobre la familia española*, de Enrique Gómez Arboleya y Salustiano del Campo, de 1957. La muestra la integran 127 estudiantes en Madrid y 112 en Salamanca. En 1959 se publica con el título *Para una sociología de la familia española*.
- *La encuesta sobre la profesión médica* en España, de José Mariano López Cepero, Amando de Miguel, Luis González Seara y José Castillo (con el asesoramiento de Juan J. Linz). La muestra la componen 500 médicos, una muestra de élites y un panel de estudiantes de Medicina.
- *La encuesta a empresarios* de Juan Linz (con la colaboración de Amando de Miguel), de 1959. Financiada por la Escuela de Organización Industrial de Madrid. La muestra la forman 460 empresarios. Algunos de los resultados principales del estudio se publican en 1966 con el título *Los empresarios ante el poder público*.
- *La encuesta a amas de casa*, de José Castillo, de 1962, relativa a hábitos de consumo.
- La primera *encuesta de juventud*, de 1960 (aunque el período de investigación se extiende de 1958 a 1965). Dirigida por José Mariano López-Cepero, un colaborador de Enrique Gómez Arboleya. Contó con un amplio equipo investigador, entre quienes se encuentra el propio Gómez Arboleya (que se incorpora tardíamente), José Bugeda, José Castillo, Luis González Seara, Juan J. Linz y Amando de Miguel. Este último autor (A. de Miguel, 2000) detalla la relación de participantes en la investigación, sus avatares y los resultados principales del estudio en un informe de la juventud comparativo: *Dos generaciones de jóvenes (1960-1998)*. Y su hermano, Jesús de Miguel (1999: 188) destaca esta encuesta, junto a las realizadas por una parte de su equipo investigador (las encuestas a médicos y

empresarios antes citadas), “el punto de arranque fundamental del análisis “empírico” de la realidad social española, con nuevos métodos importados en esa época fundamentalmente de Estados Unidos, Francia y Gran Bretaña. Aunque las tres encuestas se llevan a cabo desde Madrid”.

En 1963 se crea, a instancia de Manuel Fraga Iribarne, el *Instituto de la Opinión Pública* (que en 1976 se transforma en el actual *Centro de Investigaciones Sociológicas*). Un centro oficial de investigación sociológica empírica, con clara hegemonía de la investigación mediante encuesta (como se ha destacado en el capítulo 4). La *Revista Española de la Opinión Pública* se convierte en un medio de difusión de referencia de la investigación en España. Con posterioridad pasaría a denominarse *Revista Española de Investigaciones Sociológicas* (REIS).

En la esfera privada comienzan a aparecer institutos de investigación de mercados y de sociología aplicada. En 1958 se crea el *Instituto ECO*, vinculado a Jesús Ibáñez. En 1965 surge *DATA*, unido a las figuras de Amando de Miguel y Juan J. Linz. Bajo la dirección de Amando de Miguel se realiza el primer *informe FOESSA*, en 1966, y el segundo, en 1970. Ambos informes serían considerados “la prueba de madurez de la sociología española” (Lamo de Espinosa, 1992: 127-128). Y de manera especial, el *II Informe FOESSA* (1970) “se convierte en el mejor de los estudios sociológicos globales y, sin duda, el estudio sociológico sobre la estructura social española más avanzado para la época” (J. de Miguel, 1999: 195). No obstante, padeció la censura de la etapa final del franquismo. No llegó a publicarse el capítulo quinto sobre “vida política y asociativa”. Este capítulo se publica, con posterioridad, en el libro de Amando de Miguel (2003) sobre el final del franquismo, que incluye fragmentos de su diario personal de su paso por la cárcel en los últimos años de la dictadura.

La serie de informes FOESSA prosigue en años posteriores, pero bajo la dirección de otros investigadores y sin una periodicidad concreta. En el tercero, la encuesta se realiza en 1973, aunque el informe se publica en 1976. El cuarto, entre 1981-1983, dirigido por Juan Linz. El quinto, de mayo a junio de 1995, bajo la dirección conjunta de Miguel Juárez.

Con la instauración de la democracia, la encuesta paulatinamente va adquiriendo un mayor protagonismo en la investigación social empírica en España. En 1979, año de elecciones, en España se gastaron 2.000 millones de pesetas en encuestas de opinión (según estimaciones de López Pintor, 1982). Veinte años después, la facturación de estudios cuantitativos asciende a 18.953 millones de pesetas. En el año 2000 a 22.882 millones (Alós, 2001).

5.2. Características esenciales de la encuesta y su articulación con otras estrategias de investigación

La encuesta permite el estudio de un amplio abanico de cuestiones, tanto sobre aspectos *objetivos* (de hechos) como *subjetivos*, no directamente observables (actitu-

des, valores, opiniones), del presente, pasado y futuro. La información se obtiene mediante observación “indirecta” a través de las respuestas orales o escritas de los encuestados a las preguntas incluidas en el cuestionario, por lo que siempre cabe la posibilidad de que la información recabada no sea un fiel reflejo de la realidad que se analice. De ahí la exigencia de comprobar “siempre” la veracidad de los datos, antes de proceder a su interpretación. Además del fiel cumplimiento de sus dos características esenciales:

1. *Estandarización* en la recogida y el análisis de la información. Las respuestas se agrupan y cuantifican mediante técnicas de análisis estadístico, en busca de relaciones entre ellas. Pero, para que las respuestas puedan compararse, ha de seguirse un procedimiento estandarizado de recogida y análisis de los datos. Éstos se obtienen de forma estructurada, a partir de unas mismas preguntas, formuladas en iguales términos, y en el mismo orden, a cada uno de los encuestados. Si en la recogida de la información median entrevistadores, éstos han de seguir el mismo procedimiento de entrevista en cada encuesta, evitando cualquier actuación que introduzca desigualdad en el proceso de pregunta-respuesta. Ello posibilita la comparación de las respuestas, además de reducir errores de medición.

El logro de una medición “científica” de los fenómenos sociales precisa de la *estandarización*: las diferencias en las respuestas han de deberse a divergencias entre los encuestados, y no a desigualdades en la aplicación de la encuesta en cualquiera de sus fases.

2. *Representatividad* de la *muestra* y de la *información* respecto al objeto de estudio. La *información* ha de tener correspondencia con los valores “verdaderos” de la población; en caso contrario, habrá errores de “observación o medición”. Y la *muestra* de individuos que se observe ha de ser “representativa” de la *población* o universo de estudio. El interés del investigador va más allá del individuo concreto del que obtiene la información. Se extiende a la *población* a la que pertenece. De ahí la exigencia de representatividad de la *muestra*. Revierte en la reducción de errores de “no observación”, con el consiguiente aumento de la *validez externa* (o posibilidad de generalización de los resultados de la encuesta) y de *conclusión estadística* (o significatividad estadística de los resultados) del estudio. Como se vio en el capítulo 1, ambas se hallan afectadas por el tamaño de la *muestra* y el procedimiento de selección de las unidades que la componen.

5.2.1. Posibilidades y límites de la encuesta

El protagonismo de la encuesta en la investigación social cuantitativa se debe a las amplias *posibilidades* que ofrece al investigador. Comparada con otras estrategias de investigación, la *encuesta* permite:

- a) Obtener *información variada* de un conjunto amplio de personas, ubicadas en distintas áreas geográficas, en un plazo de tiempo relativamente breve. Éste dependerá del método de encuesta que se aplique (como se verá en el apartado 5.4). Aunque su coste sea elevado (mayor en la encuesta “cara a cara”), es inferior al que precisaría la materialización de otras estrategias de investigación para la consecución del mismo volumen de información.
- b) *Comparar* datos de encuestas realizadas en fechas, países o áreas geográficas diferentes. Ello está supeditado a la “comparabilidad” de las encuestas que se analicen, en su diseño (muestra, cuestionario) y aplicación. Como se vio en el capítulo 4, los bancos de datos de encuesta son de gran utilidad para la realización de estudios comparativos y longitudinales de tendencias.
- c) *Generalizar* sus resultados, dentro de los límites marcados en el diseño y en cada fase de la ejecución de la investigación.
- d) *Evaluar* la *calidad* de los datos, el cumplimiento de los criterios de *validez* y de *fiabilidad*. Dependerá de la magnitud y variedad de los *errores* que se detecten en su ejecución.

Pero la encuesta se enfrenta, a su vez, a una serie de obstáculos que impiden su autosuficiencia como estrategia de investigación. Entre los *límites* comúnmente señalados están:

- a) Su inadecuación en poblaciones con dificultades para la comunicación verbal (niños pequeños, personas analfabetas o con escasos recursos lingüísticos), al obtener la información de declaraciones verbales.
- b) La información está condicionada a las preguntas que incluya la encuesta, su formulación y ubicación en el cuestionario. También, a la veracidad de las respuestas dadas por los encuestados.
- c) La mediación del entrevistador (en encuestas mediante *entrevista* personal y telefónica) puede provocar efectos de carácter reactivo en la respuesta. No obstante, la *reactividad* puede reducirse con un buen diseño del cuestionario, junto con una adecuada selección y formación de los entrevistadores.
- d) Al predominar (en la investigación social usual) las encuestas *transversales* (la recogida de información se produce en un único momento en el tiempo) sobre las *encuestas panel* (o encuestación repetida de una misma *muestra* en tiempos diferentes), y carecer de un control experimental a priori, pueden confundirse simples correlaciones entre variables con verdaderas relaciones causales. Si bien, esta inexactitud puede, en parte, reducirse mediante el uso de técnicas analíticas multivariantes (control a posteriori).
- e) La realización de una encuesta precisa de la organización de un *trabajo de campo* (para la recogida de información) complejo y costoso. De manera especial en encuestas mediante entrevista personal en poblaciones amplias y dispersas.

El cuadro 5.1 resume las ventajas e inconvenientes principales de la encuesta como estrategia de investigación social.

CUADRO 5.1. Ventajas e inconvenientes de la encuesta

VENTAJAS	INCONVENIENTES
<ul style="list-style-type: none"> – Permite abarcar un amplio abanico de cuestiones en un mismo estudio. – Facilita la comparación de resultados. – Generalización de los resultados dentro de los límites marcados por el diseño y ejecución de la investigación. – Ventaja económica: obtención de un volumen importante de información a un relativo bajo coste económico y temporal. 	<ul style="list-style-type: none"> – No recomendable en poblaciones con dificultades para la comunicación verbal. – La información se restringe a la dada por el individuo (generalmente a preguntas <i>cerradas</i>). – La mediación del entrevistador puede provocar <i>reactividad</i> en la respuesta. – La carencia de referencias contextuales y vitales de los encuestados limita la interpretación de los resultados. – Imprecisión para el estudio de la causalidad. – El desarrollo de una encuesta amplia resulta complejo y costoso (más en encuestas <i>cara a cara</i>).

5.2.2. La articulación de la encuesta con otras estrategias de investigación

Algunos de los límites de la encuesta pueden solventarse con el uso complementario de otras estrategias de investigación. Como se viera en el capítulo 1, su conjunción con métodos cualitativos (en especial, las entrevistas abiertas y los grupos de discusión) contribuyen a la mejora de la encuesta tanto en la fase *exploratoria* de la investigación (la elaboración del *marco teórico*, el diseño de la muestra y del cuestionario), como de *verificación* (validación e interpretación de los resultados de la encuesta). Beneficios ya apuntados por Sieber (1973), y que pueden resumirse en:

- a) La *estructura teórica* de la encuesta puede derivarse, total o parcialmente, de hallazgos de indagaciones cualitativas. Ello precisa que, en el diseño de la investigación, los métodos cualitativos precedan a la encuesta.
- b) Los *resultados* de la encuesta pueden ser *explicados* y *validados* mediante materiales cualitativos (observaciones, entrevistas informales, grupos de discusión). “La investigación cualitativa de manera más general puede ayudar a interpretar, ilustrar y cualificar las relaciones determinadas estadísticamente” (Walker, 1988: 22). Especial relevancia adquieren cuando los resultados de la encuesta son inesperados (o sorprendentes), o cuando la encuesta ha recabado un bajo porcentaje de respuesta. Para que la indagación cualitativa

cumpla la función de *validación* de los resultados de la encuesta, se ha de realizar con posterioridad a la encuesta.

- c) Los métodos cualitativos también pueden contribuir al diseño de la *muestra*, a la identificación de sujetos de interés para ser encuestados; sujetos especiales que se descubren a partir de los hallazgos de las indagaciones de carácter cualitativo.
- d) Mejora del diseño del *cuestionario*: qué indicadores incluir y su traducción en preguntas de encuesta. Cómo enunciarlas para que sean comprendidas por la población objeto de estudio: el lenguaje a emplear para facilitar la comunicación con el encuestado. Como diría Ibáñez (1989), ha de evitarse que la *encuesta* se convierta en la opinión de quien redacta el cuestionario. De ahí su recomendación de realizar una “investigación estructural” (normalmente, algún *grupo de discusión*) como fase previa a la realización de una encuesta. El objetivo es evitar que la *encuesta* se convierta en la opinión de quien redacta el cuestionario. Por su parte, Fuller *et al.* (1993) destacan dos contribuciones claves del *grupo de discusión* al diseño del cuestionario:

- La identificación de los *conceptos* teóricos relevantes.
- La formulación de *hipótesis*.

Su consecución precisa que la encuesta se diseñe después de haberse aplicado métodos cualitativos. Si fuese a la inversa, la indagación cualitativa ayudaría al diseño del cuestionario (en su fase de validación), clarificando las preguntas que hayan quedado confusas.

La *encuesta* puede igualmente contribuir a la mejora de la investigación cualitativa. Sieber (1973) resumía en cuatro las contribuciones de la encuesta al *trabajo de campo* cualitativo:

1. *Identificar* los casos relevantes e irrelevantes para que con posterioridad sean analizados mediante indagación cualitativa.
2. *Demostrar* la generalidad de las observaciones de campo.
3. *Verificar* las interpretaciones de campo.
4. *Aclarar* las observaciones “mal interpretadas” o “inexplicables”. *Por ejemplo*, proporcionando información de informantes o individuos que fueron, previamente, pasados por alto.

El *uso de datos disponibles* también contribuye a la mejora de la encuesta, en las fases previas (elaboración del *marco teórico*, operacionalización del problema de estudio) y en las posteriores (interpretación de sus resultados). Además, son imprescindibles en los estudios comparativos, tanto transversales como longitudinales. Pero, como ya se señalara en el capítulo 1, la articulación de estrategias y técnicas de investigación social precisa que realmente se estén midiendo los mismos conceptos

para que los resultados sean comparables. Si la finalidad no es la exploración, sino la comprobación de la *validez* de los resultados, ha de garantizarse la independencia en la aplicación de las diferentes técnicas de recogida de información para evitar cualquier contagio entre sus resultados. Para más información reléase el capítulo 1.

EJEMPLO DE ARTICULACIÓN DE LA ENCUESTA CON OTRAS ESTRATEGIAS DE INVESTIGACIÓN

El proyecto *MEXEES I*, acrónimo de un proyecto I + D financiado por el MEC (SEJ2005-00568), titulado *La medición de la xenofobia en la España de comienzos del siglo XXI: nuevos indicadores y diseños de encuesta para las políticas de integración social de los inmigrantes*. Investigación dirigida por M.^a Á. Cea D’Ancona, en colaboración con Miguel S. Valles Martínez, y con la participación de Javier Álvarez Gálvez (en calidad de becario en formación). Se desarrolló en el trienio 2006-2008.

Una investigación que pretende avanzar en el conocimiento de la *xenofobia*, en sus distintas manifestaciones y medición, adecuándola al contexto sociohistórico de comienzos del siglo XXI. En investigaciones anteriores se habían constatado graves vacíos y deficiencias en la conceptualización y medición de la *xenofobia* mediante encuesta (Cea D’Ancona, 2002b, 2004). Entre los objetivos específicos de la investigación están:

1. Desarrollar nuevos indicadores (objetivos y subjetivos) que permitan radiografiar tanto la *xenofobia* “manifiesta” como la “latente”. Ello requiere el contraste de datos objetivos de *discriminación* (en ámbitos fundamentales como el empleo, la educación, la vivienda, el ocio) con subjetivos, además de otros componentes que definen las actitudes (prejuicios, estereotipos). Todo lo cual comparando la perspectiva del autóctono con la del inmigrante.
2. Analizar los factores que propician la *xenofobia*: su exteriorización y ocultación. Unos relacionados con el número y el perfil sociodemográfico de la población inmigrante. Otros, con las características socioeconómicas de la sociedad receptora, junto con los valores que en ella se defienden. Sin descuidar la influencia de los medios: las noticias sobre inmigración que se divulgan y el tratamiento que se le da (contribuyendo a crear un estado de opinión).
3. Averiguar sobre qué ejes se articula el discurso xenófobo actual y su consonancia o disonancia con las teorías existentes.
4. Comprobar los diseños de encuesta que más propician la declaración de actitudes xenófobas o contrarias a la inmigración. Quería constatarse si el uso de hojas de respuesta autorellenadas en el curso de una entrevista *cara a cara* favorece su declaración, como defiende Krysan (1998) en su estudio de las actitudes racistas o, en general, de cualquier conducta, actitud u opinión que suscite el sesgo de *de-seabilidad social* (Sudman y Bradburn, 1974, 1982/1987; Schwartz *et al.*, 1991). Además de experimentar con diferentes enunciados de preguntas, formatos de respuesta y su ubicación en el cuestionario.

Como toda investigación que busque generar un nuevo instrumento de medición que cumpla requisitos de *validez* y de *fiabilidad*, el estudio se articula en varias fases. Cada una comprende estrategias metodológicas distintas, aunque complementarias y necesarias. Una *investigación multimétodo*, que combina tres estrategias metodológicas básicas para alcanzar los objetivos del estudio: el uso de fuentes documentales y estadísticas, la indagación cualitativa mediante grupos de discusión y entrevistas abiertas, y la encuesta.

La *primera fase* incluye el uso de *fuentes documentales y estadísticas*, además de la *revisión bibliográfica*. La atención se centra en estudios teórico-empíricos sobre inmigración-racismo-xenofobia y sus distintas manifestaciones: prejuicios, estereotipos, discriminación. El énfasis se pone en la conceptualización teórica y en aspectos metodológicos relacionados con la medición de la *xenofobia* y de su opuesto, la *xenofilia* y la *integración* del inmigrante. Quería establecerse lo que el concepto *xenofobia* incluye y excluye, y su diferencia con otros conceptos relacionados. Ello se suma al análisis de *datos secundarios* de estadísticas, noticias de prensa y de encuestas archivadas en bancos de datos. Esto último imprescindible para el posterior diseño de la encuesta. El acceso a cuestionarios ya “validados” permite analizar los resultados que se obtuvieron con determinadas preguntas. Ayuda a la formulación de nuevas preguntas y a evitar errores cometidos con anterioridad.

Ésta es una fase fundamental para el encuadre de la investigación (la conceptualización del objeto de estudio y su operacionalización: la configuración del *marco teórico*) y para la interpretación de los resultados. Persigue la síntesis del estado de conocimiento hasta el presente, a nivel nacional e internacional, trazando las principales líneas de investigación desde una perspectiva histórico-comparativa.

La *segunda fase* comprende la *investigación cualitativa mediante grupos de discusión y entrevistas en profundidad* a los protagonistas (autóctonos y foráneos). Tradicionalmente la metodología cualitativa ha propiciado la captación de discursos más xenófobos o contrarios a la inmigración que la metodología de encuesta. Precisamente, porque se da más tiempo y flexibilidad a las personas para expresar sus ideas. Más si cabe, cuando se establece empatía entre el entrevistador y el entrevistado y éste le transmite confianza para expresarse con libertad; o bien, cuando el espacio de conversación es grupal y entre pares sociológicos. Se considera necesaria para diseñar una nueva encuesta.

Se articulan entrevistas *biográficas-focalizadas* (104 entrevistas: 56 españoles y 48 extranjeros) y *grupos de discusión* (9: 6 a autóctonos y 3 a extranjeros). La estructura general de la *entrevista* incluye dos amplias partes: la primera, más *autobiográfica* y libre del autóctono y del extranjero; la segunda, más *focalizada* en los asuntos conversacionales sobre la inmigración actual en España y el pasado emigrante. En ambas partes se recurre al esquema temporal convencional (pasado, presente y futuro), que se juzga metodológicamente conveniente porque permite la aproximación a la historia personal desde la que percibe y habla cada entrevistado. En los *grupos de discusión*, en cambio, no se siguió ningún guion preestablecido. Interesaba lo que primero fluye cuando se menciona el tema a hablar: la *inmigración*. Se buscaba la primera mención, pero también el orden de temas tratados y su reiteración; es decir, los aspectos que más se enfatizan, aquellos que logran focalizar la conversación grupal, sin descuidar las matizaciones que se introducen en los relatos. El proceso y resultados completos de la investigación cualitativa se detallan en Cea D’Ancona y Valles (2010a).

La *tercera fase* abarca el diseño y ejecución de la *encuesta* experimental, a partir de la información recogida en las fases precedentes. Se descarta la encuesta *telefónica*, por

su no adecuación al tema de estudio. No así los otros métodos de encuesta: la *auto-cumplimentada*, la *cara a cara* convencional y la *cara a cara con hoja de respuesta rellenada* por el propio encuestado. La investigación incluye un experimento *split-ballot*, de gran utilidad en la comprobación de los errores *de medición*. El *trabajo de campo* se desarrolló en la primavera de 2008. La *muestra* final la componen 683 españoles de 18 y más años, distribuidos aleatoriamente en las tres situaciones experimentales. En consonancia con las hipótesis de partida, la investigación mostró la existencia de efectos debidos al sesgo de *deseabilidad social* y al método de encuesta. La autoadministración del cuestionario anima la declaración de xenofobia, aunque más cuando se utilizan escalas sutiles o indirectas de rechazo versus aceptación de inmigrantes. El inconveniente está en la subrepresentación de respondientes con bajo nivel educativo en los métodos autocumplimentados (Cea D’Ancona, 2012).

El cuadro 5.2 resume ventajas e inconvenientes de la *articulación* de la encuesta con otras estrategias de investigación.

CUADRO 5.2. Ventajas e inconvenientes de la articulación de la encuesta con otras estrategias de investigación social

VENTAJAS	INCONVENIENTES
<ul style="list-style-type: none"> – Obtención de una información más profunda y diversa, al cubrirse más dimensiones de la realidad social. – Mejora la configuración del <i>marco teórico</i> de la investigación y su operacionalización. – Enriquece el diseño del cuestionario. – Ayuda a la identificación de sujetos de interés para ser encuestados. – Amplía la interpretación y <i>validación</i> de los datos de encuesta. 	<ul style="list-style-type: none"> – Un mayor coste económico y temporal. – Mayor dificultad para replicar la investigación. – Precisa de especialistas en diferentes metodologías. – La aplicación de una técnica de recogida de información puede <i>sesgar</i> los datos que recaben técnicas posteriores ⁽¹⁾. – Problemas de <i>comparabilidad</i> de los resultados, al verse afectados por diferentes fuentes de error. – La incompatibilidad metodológica de algunos métodos ⁽²⁾.

⁽¹⁾ Bien porque la experiencia alcanzada con el primer método puede llevar al investigador a utilizar el segundo de forma diferente, e inclusive mejor o bien debido a la influencia en los sujetos.

⁽²⁾ La principal objeción a la combinación metodológica con frecuencia se expresa en términos de “pertinencia epistemológica”. Para Blaikie (1991) ésta es la barrera fundamental para integrar formas de conocimiento que califica “incompatibles”.

5.3. Fases en el desarrollo de la encuesta y errores en su aplicación

La encuesta, como cualquier proceso de investigación, parte de la fase inicial correspondiente a la *formulación del problema* de investigación. A veces, es la propia entidad o institución que financia la encuesta la que propone el *problema* o tema de estudio; en otras ocasiones, es el propio investigador la que lo sugiere, en respuesta a sus inquietudes intelectuales o a intereses atribuidos a quien se espera financie la encuesta. Sea como fuere, la idea o *problema* de investigación se presenta, al principio, vaga y genérica. El investigador deberá, después, precisarla y configurarla en un diseño concreto de la investigación. Para lo cual tendrá que adentrarse en el campo de conocimiento donde se ubique el *problema* en cuestión (como ya se indicó en el capítulo 1). A tal fin, una actuación exigida es la *revisión bibliográfica* sobre el tema específico de estudio y otros relacionados. A partir de ella se podrá:

- a) Perfilar la *idea* originaria de la encuesta y confeccionar el *marco teórico* de la investigación (integrado por una o varias teorías, o por *generalizaciones empíricas* que aún no han adquirido forma de teoría). Actuará de referencia para la concreción de los *objetivos* e *hipótesis* de la encuesta, así como en la interpretación de los resultados.
- b) *Diseñar la encuesta*, teniendo presente la experiencia de encuestas similares para:
 - Evitar reincidir en los mismos *errores*.
 - No redundar en aspectos ya abordados, con objeto de seguir avanzando en el conocimiento del tema de estudio. Excepto que quiera realizarse un estudio *comparativo*. Éste precisa de la replicación exacta del estudio anterior (con el que se quiere comparar los resultados).
 - Extraer *ideas* para diseñar la *muestra* y el *cuestionario*. Como se verá en el capítulo 6, antes de proceder a redactar las preguntas del cuestionario es recomendable revisar encuestas sobre el mismo tema de estudio.

A la revisión bibliográfica se suma la consulta a *expertos* e indagación cualitativa de carácter exploratorio:

- a) Los colaboradores o *expertos* pueden aportar ideas de qué aspectos tratar y cómo llevar a cabo la investigación.
- b) Las *entrevistas* (individuales o en grupo) a sus *protagonistas* (jóvenes, jubilados, amas de casa, inmigrantes..., cualquiera que sea la población de estudio) pueden dar información que ayude al *diseño* del *cuestionario* y de la *muestra* (como se vio en el apartado anterior). Incluso pueden aportar datos discordantes con preconcepciones del investigador. En palabras de Sheatsley (1983: 203), “Los investigadores pueden encontrar que algunas consideraciones que pensaban eran importantes son irrelevantes para la gente con la que habla”.

De lo exhaustivo que se sea en esta fase inicial de la investigación dependerá el buen resultado de la encuesta.

A la concreción del *marco teórico* y objetivos de la encuesta sigue su *operacionalización*. Como puede verse en la figura 5.1, incluye: 1) la formulación de *hipótesis*; 2) la operacionalización de *conceptos* teóricos clave; 3) la delimitación de la *población* de estudio (jóvenes, amas de casa, jubilados, profesores universitarios). Esta última afecta a todo el *diseño de la encuesta*: desde la elección del método de encuesta a aplicar, el diseño de la muestra y del cuestionario, hasta la administración del cuestionario (días y horas para la realización del *trabajo de campo*, cuando la encuesta es mediante entrevista personal o telefónica).

La elección del *método de encuesta* (mediante entrevista personal, telefónica, por correo u otra forma autoadministrada) está afectada por los *objetivos* de la investigación y la *población* a estudiar. También, por los *recursos* económicos y el plazo de *tiempo* que se disponga para su ejecución. Además de la continuidad que quiera darse a la encuesta. Estos son los tres elementos claves que marcan cualquier proceso de investigación:

1. Los *objetivos* del estudio.
2. Los *recursos* económicos, humanos, materiales, que la hagan factible.
3. El *tiempo* concedido para su realización.

Su conjunción en la encuesta determinará el tamaño de la *muestra* y el procedimiento a seguir para su selección (capítulo 7). Pero también las otras fases de la encuesta, que incluye: la elección del *método de encuesta* (detallados en apartado 5.4), el *diseño del cuestionario* (capítulo 6), la ejecución y supervisión del *trabajo de campo* (capítulo 7), y el *análisis* de la información (capítulo 8). Así lo muestra la figura 5.1, que resume las fases esenciales de la encuesta.

El *método de encuesta* que se elija asimismo determinará el diseño de la *muestra* y del *cuestionario*. De la *muestra* se exige que represente a la *población* de estudio; del *cuestionario*, que se adecue a los *objetivos* de la investigación, a las peculiaridades de la población a encuestar (nivel educativo, conocimiento del tema, interés) y a los *conceptos* a medir. Éstos han de traducirse en preguntas apropiadas, que sean fáciles de comprender por las personas a quienes la encuesta se dirige. Una vez confeccionado, el cuestionario será chequeado o *pretestado*. El objetivo es comprobar su adecuación antes de pasarse a la *muestra* del estudio. También puede pasarse previamente a expertos para que evalúen su adecuación a los objetivos de la encuesta y a la medición de los conceptos principales.

La *precomprobación* del cuestionario puede llevar a fases previas del diseño de la encuesta (revisión de *objetivos*, *muestra*, *cuestionario*), además de proporcionar información para las fases posteriores (administración del *cuestionario*, *codificación*, *grabación* y *análisis de los datos*), como se verá en el capítulo 7.

Tras la redacción definitiva final del cuestionario, procede su administración a la *muestra* diseñada para la encuesta. Comienza la fase de recogida de información

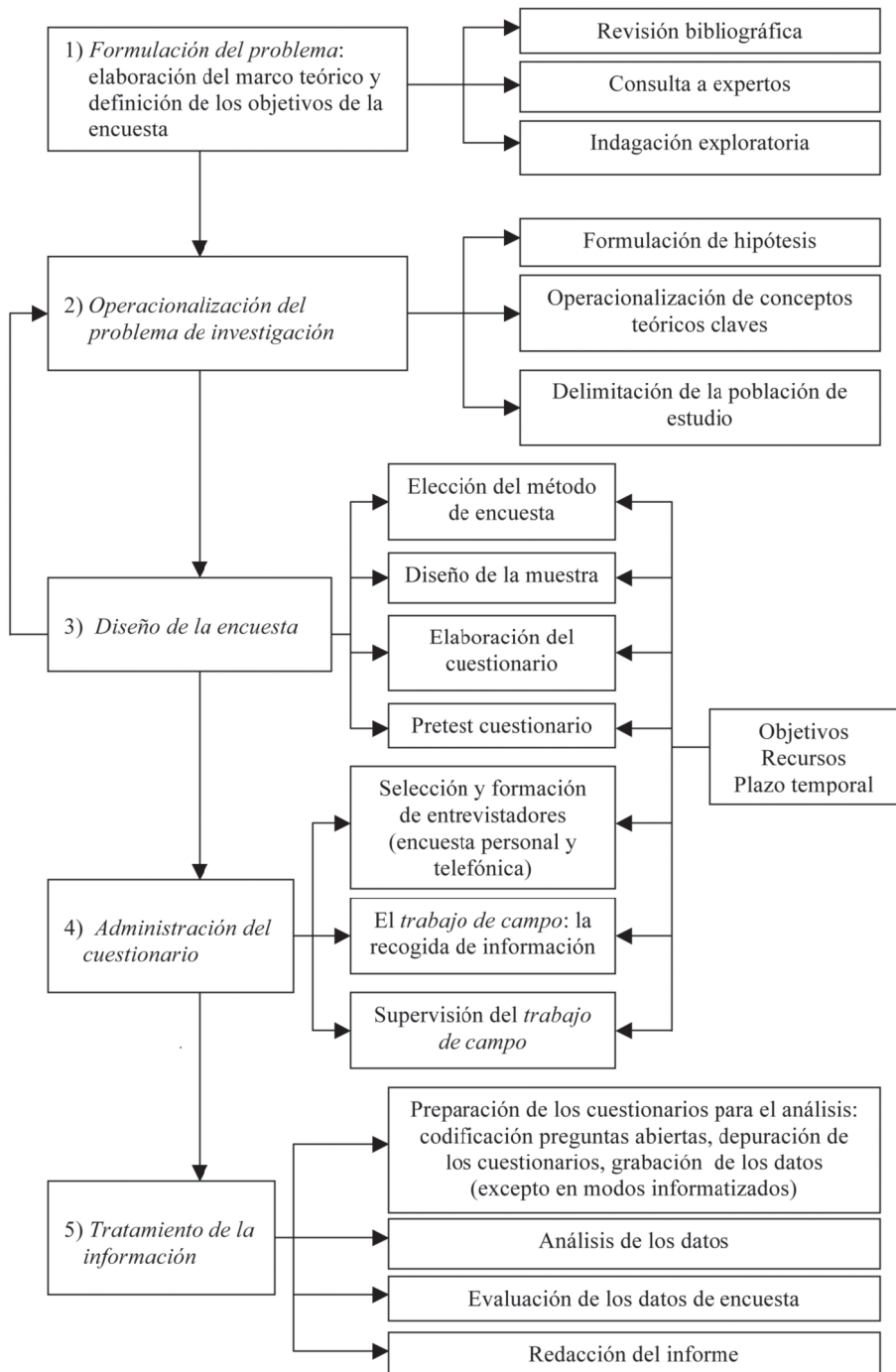


FIGURA 5.1. Fases esenciales de la encuesta.

(llamada *trabajo de campo*); la más delicada y costosa en el desarrollo de la encuesta. Su coste aumenta dependiendo del método de encuesta que se aplique, la complejidad del cuestionario y la amplitud y peculiaridades de la *muestra* que lo cumplimentará.

Cuando media la presencia de *entrevistadores*, previamente ha de procederse a su selección y formación específica para dicha encuesta. El *trabajo de campo* también tendrá que ser *supervisado*, bien in situ (encuestas telefónicas), o bien después de rellenarse los cuestionarios, para evitar errores que deterioren la *calidad* de la información recabada.

Concluida la fase de administración del cuestionario, procede la correspondiente al *tratamiento de la información*. Comprende la fase de *edición* y de *procesamiento* de los datos. Se procede al cierre de las preguntas *abiertas* (si las hubiere), al chequeo y *depuración* de los cuestionarios, y la *grabación* de los datos (salvo que se haya empleado un modo informatizado para la recogida de información).

Tras la grabación y depuración de los datos se procede a su *análisis* estadístico. Primero de carácter exploratorio, mediante gráficos y estadísticos univariantes. Después, bivariable, generalmente a través de *tablas de contingencias* o análisis cruzado de dos variables (cuyos efectos pueden a su vez controlarse por una tercera e incluso cuarta variable de *control*) y análisis de correlaciones y covarianzas. Cada vez es más imperioso que la fase de análisis culmine con la aplicación de alguna técnica analítica multivariable (en busca de relaciones causales, tipologías o agrupaciones de casos y variables como se verá en el capítulo 8).

La encuesta concluye con la redacción del *informe*. Éste cumple la función de presentación, de comunicación, de la investigación materializada. Se espera que sea una exposición pormenorizada de la encuesta en cada una de sus fases, y del alcance de sus resultados; que sea “honesto” sobre los hallazgos del estudio y que ofrezca el suficiente detalle que posibilite su replicación (capítulo 8). La especificación de los “límites” de la investigación se estima crucial para la interpretación de la información aportada y para su credibilidad.

5.3.1. Tipología de errores de encuesta

Ninguna encuesta está libre de *errores* en su cumplimentación. En cada una de las fases que comprende pueden darse uno o varios errores que mermen la *calidad* de la encuesta. Éstos se resumen en la figura 5.2. En ella aparecen agrupados dentro de una amplia categoría de errores de “*no observación*” y errores de “*observación*” o medición.

- a) Los errores de *no observación* se deben a que la encuesta no ha recabado información de toda la variedad de unidades que componen la población de estudio. Afectan a la representatividad de la *muestra* que finalmente se analiza; a las posibilidades de inferencia de los datos que de ella se extraigan. La falta de representatividad puede deberse a de errores de:

- *Cobertura*: la idoneidad del *marco muestral* (o listado de la población de estudio) para la selección de la *muestra*. Si cubre a todos los integrantes de la población objeto de estudio. La *no cobertura* puede deberse a erróneas inclusiones u omisiones de unidades de la población, y a su no correspondencia con el objeto de estudio.
- *Muestreo*: el error más medido en la investigación mediante encuesta. Su magnitud depende del tamaño de la *muestra* y de su representatividad. Si todas las unidades de la población han tenido igual probabilidad de ser seleccionadas para la *muestra*.
- *No respuesta*: la encuesta no ha extraído información de todas las unidades que componían la *muestra* inicial del estudio, ya por su no localización, ya por su negativa a participar en la encuesta (en su totalidad o de manera parcial en preguntas concretas del cuestionario).

El efecto de cualquiera de los errores de *no observación* en el deterioro de la calidad de la encuesta dependerá de la proporción que representen los excluidos de la *muestra* final y de sus características diferenciales respecto de la población finalmente incluida. En especial, si éstas se hallan relacionadas con el problema de estudio. Estos tres errores se tratan en el capítulo 7.

b) *Errores de observación o medición*. También llamados “de respuesta” (Bradburn, 1983). Representan desviaciones de las respuestas que recaba la encuesta de los valores “verdaderos”. Pueden deberse al:

- *Método de encuesta* empleado: su pertinencia para la consecución de los objetivos de la investigación.
- *Cuestionario*: las preguntas, las opciones de respuesta y su ubicación en el cuestionario pueden afectar a la respuesta, a su *validez*. Puede haber *errores de especificación* (debidos a la no correspondencia de la pregunta con el concepto que trata de medir), pero también otro tipo de errores que se analizan en el capítulo 6.
- *Entrevistador*: afecta su forma de administrar el cuestionario (cómo formula las preguntas, registra las respuestas, sus reacciones ante dificultades del encuestado para responder las preguntas o respuestas que dé), y sus características personales (capítulo 7).
- *Encuestado*: que proporcione información incorrecta bien por falta de conocimiento sobre lo que se pregunta, o bien por una mala redacción de la pregunta (que propicie su errónea comprensión), por voluntad expresa (en busca de aprobación social: el *sesgo de deseabilidad social*), o por problemas de memoria (todo ello se aborda en el capítulo 6, donde se resumen actuaciones en el diseño y administración del cuestionario para reducir su incidencia).
- *Tratamiento de la información*: comprende la *edición*, la *codificación* de preguntas *abiertas*, la *grabación* de los datos en el ordenador, la *ponde-*

ración e imputación de las respuestas, la *tabulación* de los datos y el *modelado* estadístico. Errores que resultan de actuaciones incorrectas por parte de procesadores y de analistas de los datos (capítulo 8).

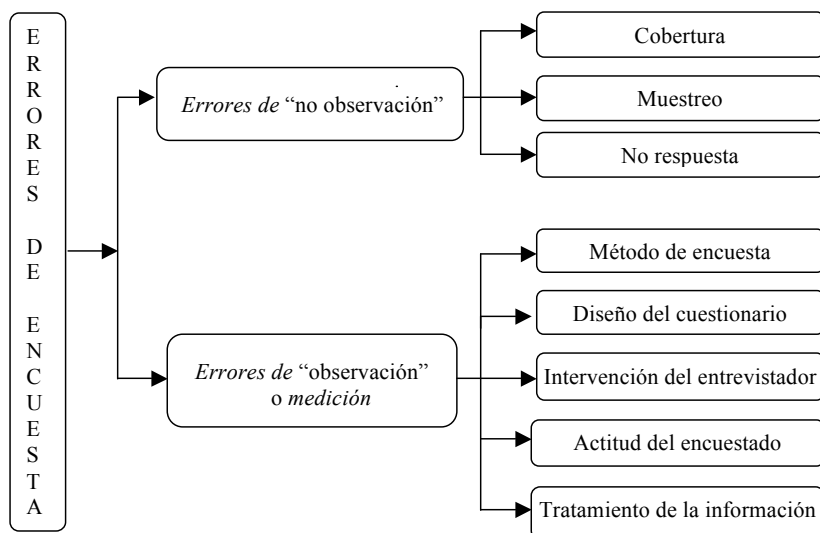


FIGURA 5.2. Errores en el desarrollo de una encuesta.

EJEMPLO DE INFORME QUE DETALLA ERRORES DE ENCUESTA

El informe realizado por SORA (Institute for Social Research and Analysis, de Viena), en marzo de 2001, para el Eurobarómetro 53 (*Attitudes towards minority groups in the European Union*), de acceso gratuito en la Red (en la página web correspondiente a los eurobarómetros). Destaca limitaciones relevantes a considerar en la interpretación de los resultados de la encuesta (que se realiza del 5 de abril al 23 de mayo de 2000 a 16.078 personas, mayores de 15 años, de los entonces quince países miembros de la Unión Europea), y que pueden llevar a cuestionar su *validez*. Afectan a:

- a) *Marcos muestrales*. Difieren bastante en calidad. En especial, lo referido a las direcciones para contactar con los encuestados. Las direcciones inválidas iban del 0% en Alemania hasta el 61% en España o el 56% en Finlandia. Porcentajes demasiado elevados que sesgaron los resultados de la investigación.
- b) *Porcentajes de respuesta*. El 62% de las 1.007 entrevistas realizadas en España fueron cumplimentadas con éxito en la primera visita, mientras que en los Países Bajos se informó de que el porcentaje de "negativas inmediatas en el primer

contacto” fue del 69%. Los porcentajes de respuesta obtenidos hasta la tercera visita (que generalmente se recomienda antes de proceder a la *sustitución* de la unidad muestral aleatoriamente seleccionada) oscilaron entre el 25% en los Países Bajos y el 81% en Francia. Dato a considerar para la interpretación comparativa de las respuestas por países. No hay certeza de que aquellos que respondieron a las preguntas del cuestionario compartiesen las mismas características de aquellos a quienes no se pudo encuestar (por no localizarlos en sus domicilios o por su negativa a participar en la investigación). Se recomienda poner más énfasis en la *estandarización* del procedimiento de selección de la *muestra* y el logro de un porcentaje mayor de respuesta.

- c) *Corrección de la muestra.* Para corregir las características no representativas de la *muestra*, se procedió a su *ponderación*. Se ajustaron desviaciones de una pequeña serie de variables como edad, género y región, en aquellos casos de los que se disponía de información detallada. Los *pesos* asignados variaron por países. Las correcciones muestrales más pequeñas se dieron en Bélgica (*pesos* entre 0,66 y 1,5) y las más elevadas en los Países Bajos (entre 0,27 y 7,29) y Suecia.
- d) Los *casos sin respuesta* difirieron bastante por países. Suecia, Dinamarca, Finlandia, Francia y Bélgica fueron los países con una menor proporción de casos *sin respuesta*. Por el contrario, en Austria, Gran Bretaña, Irlanda e Irlanda del Norte se obtuvo un elevado porcentaje de *no respuesta*. La *no respuesta* fue mayor en las mujeres de 65 años y más, jubiladas y amas de casa, en la generalidad de los países de la Unión Europea, excepto Luxemburgo.
- e) Para evitar *sesgos* debidos a la *no respuesta* (al no ser aleatoria, convergiendo con un mismo perfil sociodemográfico), se optó por la *imputación*. Se aplicó el procedimiento iterativo de maximizar lo esperado, al no estar los casos *sin respuesta* distribuidos de manera aleatoria.
- f) *Sesgos culturales* llevaron a eliminar algunas preguntas de los análisis. Aunque el cuestionario fue traducido a todos los idiomas de la Unión Europea, no pudieron evitarse diferentes connotaciones de los conceptos medido por países. Lo que propició que los autores del informe reconocieran que es un “riesgo” creer en la completa “verdad” de los resultados. En el informe especifican los países cuyos resultados no son comparables con otros países y en qué preguntas. La *comparabilidad* entre países se comprueba mediante un análisis *factorial confirmatorio* (mediante el programa LISREL), que permite comparar la estructura latente de las actitudes en todos los países. Para evitar la exclusión de los casos *sin respuesta* se utilizaron valores imputados.

El conocimiento de los *errores* habidos en el desarrollo de la encuesta sin duda ayuda a la interpretación de los resultados y a su credibilidad.

5.4. Modalidades de encuesta

Una de las decisiones iniciales, y de mayor trascendencia para el desarrollo de la investigación, es la elección del *método de encuesta* a aplicar para alcanzar los objeti-

vos de estudio. Repercute en todo el diseño de la encuesta: *muestra*, cuestionario, edición y procesamiento de la información. También, en los *errores* de encuesta (como se verá en los subapartados correspondientes). En su elección se consideran:

- a) *Objetivos* de investigación: la información que desea obtenerse. En temas sensibles, vulnerables al sesgo de deseabilidad social, los métodos autoadministrados se presentan como la mejor opción; la menos aconsejable, la encuesta telefónica. Al igual que cuando se busca una respuesta meditada o reflexiva.
 - b) *Complejidad del tema*: afecta a la extensión y a la dificultad del cuestionario. En cuestionarios complejos, con muchas preguntas *filtro*, se desaconsejan las modalidades autoadministradas. También, cuando el cuestionario incluye muchas preguntas *abiertas*. En ese caso, la encuesta *cara a cara* (mediante entrevista personal) se convierte en la mejor opción.
 - c) *Población* a encuestar: la facilidad de su contactación en fechas y momentos del día concretos, y su dispersión geográfica. Esta última puede hacer inviable la encuesta mediante entrevista personal por su mayor coste económico. Además se precisa tener información previa relativa al *marco muestral* (como se verá en el capítulo 7).
 - d) *Tiempo* disponible para la recogida de información. Dos factores que han propiciado el incremento de las encuestas telefónicas son su inmediatez y menor coste económico, en comparación con la encuesta *cara a cara* (sobre todo en el modo tradicional de lápiz y papel). La encuesta por correo requiere un mayor tiempo para la realización del *trabajo de campo* (más cuando se envían recordatorios para aumentar el porcentaje de respuesta).
- El tiempo ha de valorarse, junto con el tema de investigación. Si se prevé que su prolongación puede afectar a la respuesta, habrá que optar por la encuesta telefónica y modos informatizados (que reduzcan errores de *medición*).
- e) *Recursos* (económicos, materiales, humanos) *disponibles*. Cuando éstos son limitados, la encuesta por correo (u otra modalidad autoadministrada) se convierte en la opción más factible. La encuesta *cara a cara* continúa siendo la más costosa. Los costes afectan a todo el diseño de la encuesta: *muestra* (tamaño, heterogeneidad, dispersión geográfica), *cuestionario* (extensión, complejidad), *modo de administración* y *supervisión*. Éstos han de valorarse en relación con la calidad de los datos que se desea alcanzar. Y, como afirma Groves (1989: 130), “los errores de encuesta son inversamente proporcionales a los costes de la encuesta”.
 - f) *Nivel de calidad*. Si se antepone un mayor porcentaje de respuesta, habrá que optar por la entrevista personal. Si se prima el control del entrevistador, la encuesta telefónica. Las encuestas autoadministradas son menos vulnerables al sesgo de deseabilidad social, de gran incidencia en los *errores de medición*, pero no así de otros como se verá en el subapartado 5.4.3.

CUADRO 5.3. Métodos de encuesta, según el modo de administración

Encuestas de observación directa

El propio investigador (o personal encargado) obtiene “directamente” la información, y no a través del encuestado. El cuestionario se cumplimenta a partir de los datos extraídos de informes, expedientes u otras fuentes de información documental.

Encuestas mediante entrevistador

El entrevistador formula las preguntas y rellena el cuestionario con las respuestas dadas por el encuestado. Difieren en cómo se establece la comunicación entre el entrevistador y el entrevistado: “cara a cara” o a través del hilo telefónico.

a) Encuesta “cara a cara” o mediante entrevista personal

- PAPI = Paper-and-Pencil Personal Interviewing
- Modos informatizados:
 - CAPI = Computer-Assisted Personal Interviewing
 - HAPI = Handheld Personal Interviewing

b) Encuesta telefónica

- Modos informatizados:
 - CATI = Computer-Assisted Telephone Interviewing
 - Entrevistador “virtual”:
 - IVR = Interactive Voice Response
 - TDE = Touchtone Data Entry
 - T-ACASI = Telephone-Audio Computer Assisted Self-interview
 - VR = Voice Recognition
 - ASR = Automatic Speech Recognition

Encuestas autoadministradas

Los encuestados leen y rellenan el cuestionario sin la mediación del entrevistador.

a) Entregadas por un entrevistador

- Encuestas a grupo
- Encuestas diario
- Modos informatizados:

CUADRO 5.3. (Continuación)

-
- CASI = Computer-Assisted Self-Administered Interviewing
 - CSAQ = Computerized Self-Administered Questionnaire
 - PDE = Prepared Data Entry
 - ACASI = Audio-Computer-Assisted Self-Administered Interviewing
 - DBM = Disk by Mail

b) Envío por correo

- Encuesta por correo postal
- Encuesta por fax
- Modos informatizados:
 - DBM = Disk by Mail
 - EMS = Electronic Mail Survey
 - Encuesta web; CAWI = Computer Assisted Web Interviewing

Combinación de métodos de encuesta

En una misma investigación se articulan dos o más métodos de encuesta.

* Los modos informatizados de encuesta suelen referenciarse, de forma genérica, bajo las siglas: CADAC = *Computer-Assisted Data Collection*; CASIC = *Computer-Assisted Survey Information Collection*; CAI = *Computer-Assisted Interviewing*.

El cuadro 5.3 ofrece una clasificación de métodos de encuesta atendiendo al modo de administración del cuestionario. Incluye los modos tradicionales (lápiz y papel) y sus correspondientes aplicaciones informáticas. Los siguientes subapartados versan sobre las modalidades de encuesta de mayor aplicación.

5.4.1. Encuesta “cara a cara” o mediante entrevista presencial

Tradicionalmente asociada a la obtención de datos de “calidad”. Proporciona el máximo grado de interacción entre el entrevistador y el entrevistado, con las ventajas e inconvenientes que ello supone. Es el *entrevistador* quien formula las preguntas y anota las respuestas en el cuestionario. Razón por la cual, el éxito de la encuesta descansa, en gran medida, en la actuación del entrevistador. El cuadro 5.4 resume las ventajas e inconvenientes de la encuesta *cara a cara*. Entre las *ventajas* destacan:

- a) *Mayor versatilidad y flexibilidad de contenido.* Permite el uso de preguntas abiertas y poco estructuradas.
- b) *Más adecuada para el tratamiento de temas complejos.* El entrevistador puede aclarar cuestiones no entendidas, de forma verbal y gráfica, y es más probable que la duración de la entrevista pueda alargarse, a diferencia de la encuesta telefónica (que no ha de sobrepasar los quince minutos). El cuestionario puede ser extenso y complejo, con preguntas *filtro* (no recomendables en encuestas *autoadministradas*) y preguntas *cerradas* con muchas opciones de respuesta (no conveniente en la encuesta *telefónica*), al poderse utilizar *tarjetas de respuesta*. Éstas permiten que el entrevistado lea las opciones de respuesta, reduciendo *errores de medición* en preguntas *cerradas* con más de tres alternativas de respuesta.
- c) *Permite respuestas de mayor calidad,* pensadas con más tiempo (que la entrevista *telefónica*) y clarificadas por el entrevistador (a diferencia de las encuestas *por correo*). El entrevistador puede comprobar la comprensión de las preguntas y la consistencia de las respuestas.
- d) *Mayor cobertura de la población general y de accesibilidad a determinados grupos de población* (personas mayores, con dificultades de audición, de comprensión de preguntas, poblaciones no cubiertas por otras modalidades de encuesta como las encuestas telefónicas y web).
- e) *Menor tasa de no respuesta (total y de ítem),* al haber contacto visual entre encuestado y entrevistador, pudiendo éste aclararle cualquier duda y motivarle para responder (con sinceridad y claridad). La encuesta *cara a cara* además registra una menor *no respuesta* “parcial”: una vez comenzada la entrevista, es menos probable que se interrumpa (a diferencia de las encuestas telefónicas y autoadministradas). En especial, cuando se genera empatía entre entrevistador-entrevistado (En el meta-análisis de Hox y DeLeeuw [2002] las encuestas *personales* alcanzaron el mayor porcentaje de respuesta [70%]; seguida de la encuesta *telefónica* [67%] y por *correo* [61%]).
- f) *Mayor certeza que el cuestionario lo cumplimenta la persona elegida para la muestra* (a diferencia de las encuestas *telefónica* y *autocumplimentada*).
- g) *El entrevistador puede recabar información* (“*subjetiva*”) *complementaria del entrevistado,* ajena al cuestionario: sus características personales (aspecto físico, comportamiento durante la entrevista, gestos ante preguntas que lleven a dudar de la sinceridad de su respuesta), su entorno (características de la vivienda, del barrio), y hechos que acontezcan durante la entrevista y afecten a la respuesta (interrupciones, distracciones).

Entre las *desventajas* principales se hallan:

- a) *Elevado coste económico y temporal* (superior a otros métodos). El coste se dispara en encuestas a poblaciones muy dispersas y con muchos puntos de muestreo (capítulo 7).

- b) *Menor control de la actuación del entrevistador* que la encuesta telefónica (caracterizada por la proximidad del supervisor durante su desarrollo).
- c) *Mayor dificultad para acceder a domicilios particulares y determinados grupos de población*, con la consiguiente merma de la representatividad de la muestra final. Personas de difícil acceso, bien porque no se hallen en sus domicilios cuando se realiza el *trabajo de campo*, bien por una excesiva vigilancia que impide el acceso de entrevistadores (estratos socioeconómicos elevados), o bien por su inexistencia (en los estratos más bajos). Esta dificultad para contactar con determinados grupos de población puede impulsar a los entrevistadores a “falsificar” la entrevista (Biemer y Lyberg, 2003). Pero, sobre todo genera errores de *no observación* que impiden la generalización de los resultados de la encuesta a la población en general.
- d) *La presencia del entrevistador puede suscitar reactividad* en la respuesta de los encuestados. De manera especial, en temas sensibles, que puedan considerarse íntimo-personales, y vulnerables a la deseabilidad social (a diferencia de las encuestas *autoadministradas*). Una buena opción puede ser el empleo de *hojas de respuesta* autoadministradas en el contexto de una entrevista *personal*, ya recomendado por Sudman y Bradburn (1974), y después corroborado en estudios sobre consumo de alcohol (Aquilino, 1994) y racismo y xenofobia (Krysan, 1998), *por ejemplo*.

CUADRO 5.4. Ventajas e inconvenientes de la encuesta cara a cara

VENTAJAS	INCONVENIENTES
<ul style="list-style-type: none"> – Mayor versatilidad y flexibilidad de contenido – Adecuada para el tratamiento de temas complejos y delicados – Amplia cobertura de la población general – Menor <i>no respuesta</i> total y de ítem – Mayor certeza de que la persona que contesta el cuestionario es realmente quien ha de responder – Permite recabar información complementaria del encuestado y su entorno – Puede controlarse la actuación del encuestado 	<ul style="list-style-type: none"> – Encarece los costes económicos y temporales del estudio – Problemas de acceso a domicilios particulares y a grupos concretos de población – La presencia del entrevistador puede provocar <i>reactividad</i> en el encuestado – Vulnerabilidad al sesgo de <i>deseabilidad social</i> – Control de la actuación del entrevistador

Algunos de los inconvenientes de la encuesta *cara a cara* pueden resolverse a través de los modos informatizados. A finales de los años ochenta, comienza la expansión de las *entrevistas personales asistidas por ordenador* (CAPI = Computer-As-

sisted Personal Interviewing). A ello contribuyó el éxito de la aplicación automatizada de la *entrevista telefónica* (CATI: *Computer-Assisted Telephone Interviewing*), a partir de los años setenta, y la mayor disponibilidad de ordenadores portátiles. CAPI nace en Holanda y después se extiende a Estados Unidos y a otros países europeos. En Estados Unidos, CAPI se aplica en las principales encuestas federales (como *National Medical Expenditure Survey*, *National Longitudinal Survey of Youth*, *Current Beneficiary Survey*, *National Survey of Family Growth*). En España, su uso es aún limitado, por la gran inversión en ordenadores portátiles que precisa. Si bien, la grabación directa de las respuestas en el ordenador acaba abaratando los costes de la investigación en tiempo y dinero.

Las *ventajas* de CAPI son extensibles a otros modos informatizados de encuesta, y se resumen en el cuadro 5.5:

- a) Al integrar tres fases de la encuesta (recogida de información, grabación de datos y edición), *se reducen los costes económicos y temporales*. La reducción de costes afecta a las partidas correspondientes a encuadernación, reprografía, grabación y edición de los datos.
- b) *Mejora la calidad de los datos*, al disminuir errores de administración del cuestionario y de grabación de las respuestas. La aplicación informática facilita la actuación del entrevistador, proporcionándole la *ruta* a seguir para la formulación de preguntas a cada entrevistado. Su efecto adquiere un relieve especial cuando el cuestionario incluye varias preguntas *filtro* (que “filtran” a encuestados concretos, como se verá en el capítulo 6). Los *saltos* de pregunta se realizan de manera automática. Es el programa quien decide, a partir de las respuestas a preguntas precedentes, qué preguntas proceden a continuación, reduciendo errores debidos a una mala realización de las preguntas *filtro* (que incrementan la *no respuesta* de ítem).
- c) *Detecta errores de registro y de consistencia de las respuestas*, que pueden solventarse en el transcurso de la entrevista. En cada pregunta se especifica tipo de pregunta, clase de respuesta (numérica, alfanumérica), amplitud de las respuestas. Cualquier respuesta no incluida en el rango de opciones posible es detectada en el mismo momento de su registro, posibilitando su reformulación, antes de concluir la entrevista.
- d) *Reduce errores por orden de las preguntas y las respuestas*, al poderse programar el cuestionario para que ofrezca a cada entrevistado un orden de preguntas o de respuestas diferente. Ello permite reducir los efectos de *recencia* (elección predominante de la última opción de respuesta), de *primacía* (la primera opción) y demás tratados en el capítulo 6.
- e) *Facilita la grabación y codificación de las preguntas abiertas*. Si se anticipan las posibles respuestas válidas, pueden reducirse errores de registro.
- f) *Mayor control de la actuación del entrevistador*, al grabarse información relativa a la ejecución de la entrevista: hora de comienzo y de finalización, intervalo entre entrevistas y la secuencia en su realización. Conocer la duración de la entre-

vista contribuye al control de la calidad de la encuesta. Permite comprobar que se ha dado al entrevistado tiempo suficiente para responder las preguntas. Además contribuye a evitar la práctica (voluntaria o no) de sólo formular algunas preguntas del cuestionario y, a partir de las respuestas, rellenar el resto.

- g) *Transmite una imagen de mayor profesionalidad* del entrevistador y del organismo que realiza la investigación. Ello afecta a la participación en la encuesta.
- h) *Menor inhibición* en preguntas sensibles o delicadas (De Leeuw, Hox y Snijders, 1995/2002), reduciendo el *sesgo de deseabilidad social*. Pero siempre que transmita privacidad de la respuesta. Escribir las respuestas directamente en el ordenador puede suscitar un mayor *anonimato* de la respuesta, al reducirse las manos por las que pasan los cuestionarios.
- i) *Incrementa las opciones de ayuda visual* para facilitar la comprensión de las preguntas. Las opciones gráficas y de tratamiento de imágenes se han ampliado notoriamente en los últimos años.
- j) *Facilita la realización de encuestas complejas*, con cuestionarios con bloques de preguntas que se repiten, cuando se solicita información específica de diversos objetos o sujetos de estudio. Cualquiera de los modos informatizados permite indicar el número de miembros de la familia (u otro objeto de interés) e, incluso, asignarle un nombre. Para cada uno de ellos aparece el bloque de preguntas a realizar. Preguntas relativas a “todos los vehículos poseídos en un hogar, acontecimientos que se están investigando en detalle, o todos los establecimientos con los que se tiene una relación laboral, se pueden tratar de esta forma, evitando las confusiones por parte de tanto los entrevistadores como de los entrevistados” (Biemer y Lyberg, 2003: 192).

Los *modos informatizados* contribuyen a mejorar la *calidad* de la encuesta. Pero, como ya indicaban Nicholls II, Baker y Martin (1997: 242), “no proporcionan una panacea para el error de no cobertura, de no respuesta y de medición”. Además, la mediación del ordenador puede asimismo generar efectos negativos en el desarrollo de la encuesta. Entre sus *inconvenientes* están:

- a) *El diseño de un cuestionario “virtual” precisa más tiempo y dinero* que el cuestionario tradicional de lápiz y papel. Estos costes varían dependiendo de la extensión y complejidad del cuestionario. Si bien, pueden amortizarse cuando la *muestra* es elevada. Esta partida se suma a la inicial de equipamiento informático y formación de personal. Pero, una vez realizada la inversión inicial, su aplicación reiterada en estudios posteriores supone un considerable ahorro en los costes globales de la encuesta. Precisamente, al conjugarse las fases de recogida de datos, grabación, depuración y edición, se convierte en una opción económica para futuras encuestas.
- b) *Exige una preparación añadida en los entrevistadores* en el manejo del programa informático. Ha de evitarse que estén continuamente mirando al teclado y la pantalla del ordenador para no alargar la duración de la entrevista.

ta, además de perder contacto visual con el entrevistado (una de las ventajas exclusivas de la encuesta *cara a cara*). Además, la pérdida de contacto visual puede inhibir al entrevistado, llevándole a acortar la respuesta (de manera especial en preguntas *abiertas*). La excepción, en temas sensibles, que inhiben, porque transmite una mayor privacidad de la respuesta.

- c) *La no familiaridad con los ordenadores puede inhibir la participación.* Si bien, este inconveniente es más habitual en encuestas *autoadministradas*.
- d) *No ver el teclado* (al grabar las respuestas) puede llevar a *errores de grabación*. Si la respuesta errónea (*por ejemplo*, grabar 43 años en vez de 34 o marcar la respuesta codificada con el número 4, cuando en realidad se dijo 3) está incluida en el rango de respuestas programadas, es difícil percatarse del error de grabación durante la entrevista. Excepto cuando se empleen *pantallas de resumen* de las preguntas del cuestionario (de gran utilidad para detectar incongruencias en las respuestas), o si el entrevistador, antes de concluir la entrevista, repasa las respuestas grabadas.

CUADRO 5.5. Ventajas e inconvenientes de CAPI y demás modos informatizados de encuesta

VENTAJAS	INCONVENIENTES
<ul style="list-style-type: none"> – Elimina las partidas de encuadernación, reprografía, grabación y edición – Reduce el tiempo para disponer de los resultados de la encuesta – Elimina errores de administración de las preguntas <i>filtro</i> y las <i>rutas</i> a través del cuestionario, reduciendo la <i>no respuesta</i> de ítem – Detección inmediata de errores de registro y consistencia de respuestas – Evita errores por orden de las preguntas y de las opciones de respuesta – Facilita la grabación y codificación de las preguntas <i>abiertas</i> – Aumenta el control de la actuación del entrevistador – Transmite una imagen de mayor profesionalidad – Contribuye a reducir el sesgo de <i>deseabilidad social</i> – Incrementa las opciones de <i>ayuda visual</i> – Favorece la realización de encuestas complejas, con bloques de preguntas diferentes para distintas unidades de observación 	<ul style="list-style-type: none"> – Mayor inversión inicial en equipamiento informático y formación de personal – El diseño del cuestionario requiere un mayor coste económico y temporal – Requiere una mayor formación de los entrevistadores – La mediación de un ordenador puede inhibir la participación – Los errores de grabación no detectados en el transcurso de la entrevista son más difíciles de resolver (que en cuestionarios impresos) – La propia limitación de la pantalla del ordenador. Su tamaño dificulta la lectura y contextualización de las preguntas del cuestionario

5.4.2. Encuesta telefónica

Aunque sus orígenes se sitúan en los años treinta (Massey, 1988), no es hasta la década de los setenta cuando comienza su expansión. A ello contribuyó la mayor cobertura telefónica y los avances informáticos, con el nacimiento de CATI (*Computer-Assisted Telephone Interviewing*), que fue utilizado por primera vez en 1971, en una investigación sobre igualdad de oportunidades realizada por Chilton Research of Radnor, en Pensilvania (EE UU). Desde entonces, la encuesta *telefónica* se vincula a CATI, que progresivamente ha ido perfeccionándose introduciendo módulos específicos para el diseño de la *muestra* y del cuestionario:

- *RDD (Random Digit Dialing)*, que automatiza la selección de la *muestra* y el marcaje y llamada de los números de teléfono.
- Los *módulos de diseño de cuestionario* ofrecen la opción de grabar las preguntas con una voz predeterminada y de traducción a varios idiomas: ACASI (*Audio-Computer-Assisted Self-Interviewing*) para encuesta telefónica y autoadministrada.
- El *soporte de las redes locales LAN (Local Area Network)*, que ayudan a la supervisión de los entrevistadores. Desde la central, y en cualquier momento, se tiene información sobre la actuación de los entrevistadores, pudiéndose transmitir cualquier tipo de mensaje (Fernández Prados [1999] detalla diferentes programas informáticos para el desarrollo de la encuesta telefónica).

El cuadro 5.6 resume las ventajas e inconvenientes específicos de la encuesta *telefónica*. Entre las *ventajas* principales están:

1. *Abarata los costes del trabajo de campo* en comparación con la encuesta *cara a cara*. Precisa menos personal (entrevistadores, supervisores) y elimina la partida presupuestaria destinada a costes de desplazamiento de entrevistadores.
2. *Acorta el tiempo de realización del trabajo de campo*. Desde un mismo local, un reducido número de entrevistadores puede, en un mismo día, efectuar un número elevado de entrevistas (imposible de alcanzar mediante la entrevista *cara a cara*). Ello permite disponer de los datos de encuesta incluso en menos de una semana, siendo el método más idóneo cuando se precisa “inmediatez”.
3. *Permite abarcar núcleos dispersos de población, sin apenas coste* (en tiempo y dinero), al no tenerse que desplazar los entrevistadores. Favorece la realización de diseños muestrales ideales (el muestreo aleatorio *simple* y el *sistemático*), no siempre factibles en encuestas *cara a cara* a poblaciones espacialmente dispersas. Ello redunda en heterogeneidad y consiguiente representatividad de la *muestra*.

4. *Puede contactarse con grupos de población menos “accesibles”* por su profesión, características de la vivienda o del entorno donde residan: personas con jornada laboral que coincide con las horas habituales para el *trabajo de campo*, que residen en edificios vigilados o que habitan en viviendas retiradas o en barrios marginales “peligrosos”.
5. *Facilita la repetición de intentos de selección de las unidades de la muestra.* Es más cómodo y económico llamar “reiteradamente” a un mismo número de teléfono que reenviar al domicilio a un entrevistador. Por lo que las *rellamadas* pueden aumentarse a más de tres (en diferentes horas y días), antes de proceder a la sustitución de la unidad de la *muestra* por otra similar.
6. *Posibilita la supervisión de la entrevista durante su realización.* La confluencia en un mismo espacio físico de entrevistadores y supervisores permite el asesoramiento continuo y un mayor control de su actuación (escuchar la entrevista).
7. *Reduce el efecto del entrevistador*, al estar su actuación más controlada y estandarizada (mayor uniformidad en la realización de la entrevista, tono de voz a utilizar). Además se elimina el efecto de aspectos externos (apariencia física, vestimenta, gestos), al no haber un contacto visual con el entrevistado.

Pese a sus potencialidades, la encuesta *telefónica* se enfrenta a *inconvenientes* que limitan su uso generalizado en la investigación social:

- a) *Error de cobertura*, aunque la cobertura telefónica ha crecido bastante en los últimos años. La encuesta telefónica excluye a las personas que carecen de teléfono o cuyo nombre no aparece en el listín telefónico (siempre y cuando éste sea el *marco* para la selección de la *muestra*, no cuando los números de teléfono se generen mediante marcaje aleatorio de dígitos, RDD: *Random-Digit-Dialing*). A ello se suma que continúa habiendo diferencias por hábitat y estatus social, en contra de los municipios rurales y personas de menor nivel de renta; lo cual merma la representatividad de la *muestra*, cuando los resultados de la encuesta se infieran al conjunto de la población (no a usuarios de teléfonos). De ahí el recurso a los ajustes estadísticos de *ponderación e imputación* (capítulo 7) para compensar los errores de *cobertura*. Su disminución dependerá de las diferencias entre quienes son entrevistados y aquellos que no lo son.
- b) *Error de no respuesta (total y de ítem)* superior a la encuesta *personal*, aunque inferior a la encuesta *por correo* (postal o electrónico). La magnitud de este error también depende de las características diferenciales entre quienes integran la *muestra* original y los finalmente entrevistados. Más si éstas están relacionadas con el problema de investigación.

En la *encuesta telefónica* la *no respuesta* puede deberse a que no logra contactarse con las unidades de la muestra (contestadores automáticos, servicio de identificación de llamada), que ésta rehúse participar o a impedi-

mentos físicos que dificulten la comunicación telefónica (problemas de audición, afonía, comprensión verbal...). Para aminorar la incidencia de la *no respuesta*, se recomiendan cuatro actuaciones factibles:

1. Incrementar las rellamadas.
2. Mantener el teléfono sonando más tiempo antes de colgarlo.
3. Acortar la presentación de la encuesta. Cuanto más extensa sea la presentación, más probable es que el entrevistado pierda interés y rehúse participar en la encuesta (Dillman, Gallegos y Frey, 1976; Couper y Hansen, 2002).
4. Motivar a la persona contactada para que cumplimente el cuestionario (destacando la relevancia de la información que pueda aportar para los objetivos de la investigación).

- c) *Limitaciones en el diseño del cuestionario*: su complejidad y extensión. Mientras en la *encuesta cara a cara* son usuales entrevistas de 30 a 40 minutos, en la *telefónica* se aconseja que no sobrepasen el cuarto de hora. La carencia de contacto visual entre entrevistador-entrevistado facilita que se interrumpa la entrevista (colgar el teléfono: *no respuesta* parcial). Cuestionarios breves, con predominio de preguntas *cerradas*, sencillas, con pocas opciones de respuesta.
- d) *No pueden usarse ayudas visuales* (tarjetas u hojas de respuesta, dibujos, imágenes) que faciliten la comprensión de la pregunta. En cuestionarios complejos una práctica de gran éxito (en encuestas del *National Opinion Research [NORC]*, de Estados Unidos) ha sido el envío de materiales visuales con antelación a la entrevista telefónica (Bradburn, 1983). Para Wert (2000: 210) el uso combinado teléfono-correo, teléfono-fax, anula las ventajas de la recogida telefónica en tiempo y coste, convirtiéndose en “recursos más bien teóricos” que prácticos.
- e) *Mayor vulnerabilidad a los sesgos de aquiescencia* (acuerdo o respuestas de conformidad), de *evasividad* (no emitir opinión: no sabe o no contesta), elegir las respuesta más extremas en la escala (la primera: *primacia*; o la última: *recencia*) (Jordan, Marcus y Reeder, 1980). Su efecto puede aminorarse con alguna de las actuaciones siguientes:
- Solicitar al entrevistado que escriba las respuestas a medida que se le van leyendo.
 - Uso de escalas sencillas y numéricas (de 0 a 10, por ejemplo).
 - Desdoblar categorías de respuesta complejas.

CUADRO 5.6. Ventajas e inconvenientes de la encuesta telefónica

VENTAJAS	INCONVENIENTES
<ul style="list-style-type: none"> – Menor coste económico y temporal para la realización del <i>trabajo de campo</i> – Precisa de menos personal que la entrevista <i>cara a cara</i> – Permite abarcar núcleos dispersos de población sin apenas coste – Puede contactarse con grupos de población menos “accesibles”, por sus características personales, de vivienda o lugar de residencia – Facilita la rellamada – Posibilita la supervisión de las entrevistas mientras se realizan – Disminuye el efecto del entrevistador, por el mayor control y estandarización de su intervención 	<ul style="list-style-type: none"> – Errores de <i>cobertura y no respuesta (total y de ítem)</i> – Barreras físicas (contestadores automáticos, líneas ocupadas, teléfonos desconectados, identificación de llamadas) impiden el contacto con unidades muestrales – Fatiga más que la encuesta <i>personal</i> y limita la complejidad y extensión del cuestionario – No puede acudir a <i>ayudas visuales</i>, salvo que se complemente con otros métodos de encuesta – Exige mayores habilidades persuasivas y de conversación (por parte del entrevistador), y de memoria (retención de preguntas y respuestas) del entrevistado – No permite recabar información suplementaria del entrevistado – Más expuesta a <i>sesgos</i> de respuesta (aquiescencia, evasividad, primacía, recencia) – Respuestas menos pensadas y más breves

En los últimos años han ido apareciendo modalidades de encuesta *telefónica* “automatizadas”, donde el entrevistador se sustituye por uno “virtual” (programado para formular preguntas, indicar opciones de respuesta y repetir las elegidas para verificar que éstas se han anotado correctamente). Difieren en el procedimiento para registrar las respuestas:

- IVR (*Interactive Voice Response*), T-ACASI (*Telephone-Audio Computer Assisted Self Interview*) y TDE (*Touch-Tone Data Entry*): formas autoadministradas de entrevista telefónica. El entrevistado registra la respuesta “pulsando” (en el teclado del teléfono) el número que corresponde a la respuesta elegida. Para ello, la voz “automatizada” ha sido programada para indicarle el número que corresponde a cada respuesta. *Por ejemplo*, 1 “sí”, 2 “no”; 1 “de acuerdo”, 2 “en desacuerdo”, 3 “indiferente”. Tras haber pulsado el número, la voz repite la respuesta para corroborarla antes de su registro definitivo. Precisa que el teléfono sea de “pulsación”. Lo que aumenta el error de *cobertura*.
- VR (*Voice Recognition*) y ASR (*Automatic Speech Recognition*): el encuestado se limita a decir la respuesta al teléfono. La voz “automática” la verifica, repitiéndola antes de su grabación final. Este proceder permite su utilización

en cualquier teléfono, aunque tiene el inconveniente principal de la interpretación correcta de la respuesta.

La aplicación de cualquiera de estos modos automatizados aún es limitada. Se adecuan a encuestas muy breves (de 5 a 10 minutos, como máximo), con cuestionarios con pocas preguntas, *cerradas*, cortas, sencillas y escasas opciones de respuesta. Del investigador se exige una mayor previsión de cualquier incidencia que pueda darse en la comprensión o grabación de las respuestas. A su favor, cinco *ventajas* principales:

1. *Abarata los costes de la investigación*, al no precisar de entrevistadores ni de supervisores que les controlen.
2. *Reduce aún más el tiempo* para la recogida de información.
3. *Amplía las horas y días* en que puede realizarse la entrevista. El servicio está abierto las veinticuatro horas del día, los siete días de la semana.
4. *Mayor control del efecto del entrevistador* en la respuesta (cómo formula las preguntas y registra las respuestas), al ser un entrevistador “virtual”.
5. *Puede llevar a una mayor sinceridad en la respuesta*, porque no se dice a una persona, sino a una “máquina”, lo que incrementa el anonimato.

En su contra, el no beneficiarse de la ventaja más reconocida de la encuesta mediante entrevista: poder aclarar cualquier duda que surja durante el desarrollo de la entrevista (excepto que se hayan previsto, programando al respecto). Esta *desventaja* principal se suma a la mayor brevedad y sencillez de la encuesta. Todo lo cual limita su aplicación a la investigación social.

5.4.3. Encuestas autoadministradas

El encuestado lee el cuestionario y anota las respuestas. Puede estar, o no, acompañado por alguna persona responsable del estudio, que pueda resolver cualquier duda sobre la encuesta y su cumplimentación. Este es el caso de las encuestas realizadas en recintos cerrados (aulas escolares, oficinas, centros comerciales, hospitales...). También están las encuestas por *correo* postal o electrónico y las encuestas *web*, que omiten la mediación de un entrevistador. El encuestado lee el cuestionario, lo rellena y envía a la dirección del equipo investigador. Ello permite:

- a) *Abaratar los costes del trabajo de campo*. Se eliminan las partidas de formación y retribución de entrevistadores y supervisores, además de las correspondientes a dietas y costes de desplazamiento. Por lo que puede aumentarse el tamaño de la *muestra*, sin apenas incrementar el coste de la encuesta.
- b) *Ampliar la cobertura del estudio*. Alcanzar áreas aisladas de población y personas difíciles de localizar en sus domicilios particulares.

- c) *Cumplimentar el cuestionario en el día y hora que el encuestado elija* (no cuando le “aborde” el entrevistador). Lo que puede contribuir a mejorar la calidad de la respuesta, porque dispone de más tiempo para reflexionarla. Además, puede realizar cualquier consulta (a otra persona o documento) que estime necesario para responder al cuestionario. Ello convierte a la encuesta por *correo* en el método más adecuado, cuando se requiere información detallada y reflexiva; y no una respuesta espontánea.
- d) *Transmitir mayor privacidad y anonimato de la respuesta*, al anotarse directamente en el cuestionario. Rasgo distintivo que convierte a la encuesta *auto-administrada* en la más idónea para el tratamiento de temas delicados, íntimo-personales, al ser menos vulnerables al sesgo de la *deseabilidad social* (Tourangeau y Smith, 1996; De Leeuw, 2008).
- e) *Formular preguntas con muchas opciones de respuesta* porque pueden visualizarse. Asimismo permite el uso de *ayudas visuales* (figuras, cuadros, imágenes, fotografías, mapas) para abordar determinadas cuestiones.

Pese a sus ventajas, la encuesta por *correo* aún es de uso limitado debido a la confluencia de una serie de *inconvenientes* principales, también resumidos en el cuadro 5.7:

- a) *Inasistencia al encuestado* para que estructure su respuesta o comprenda términos de la pregunta. Esto restringe su aplicación en personas de bajo nivel educativo.
- b) *Imposibilidad de controlar si fue la persona seleccionada* (en la muestra) *quien realmente rellena el cuestionario*. Tener acceso a esta información es trascendental para los posteriores ajustes de la *no respuesta*. Además es importante saber, para el análisis de los errores de *medición*, si la persona contesta el cuestionario sola o con ayuda de otras personas; si lo cumplimenta en un ambiente tranquilo o su atención es continuamente interrumpida. Todo ello afecta a la calidad de la respuesta.
- c) *El encuestado puede leer todo el cuestionario antes de rellenarlo*. Ello reduce la eficacia de las preguntas de *control* (para comprobar la veracidad y coherencia de las preguntas) y de cualquier acercamiento progresivo a determinadas cuestiones (técnica del *embudo*: preguntas cuyas respuestas han de seguir un orden concreto). Tampoco se recomienda el uso de preguntas *abiertas*, porque su respuesta será “más breve”, “difícil de interpretar” y, con mayor probabilidad, “pasadas por alto”, que en la encuesta mediante entrevista (Dillman, 1983: 369).
- d) *Problemas de cobertura*, dependiendo de lo actualizado y completo que sea el *marco* elegido para la selección de la *muestra*. Exige disponer de un censo “nominativo” (con nombre, apellidos, dirección). Si se adjunta el número de teléfono, mejor para posteriormente poderse utilizar para el seguimiento de la *muestra*. Cuando quiera que el listín telefónico se emplee para extraer la

muestra, la encuesta por *correo* estará afectada por iguales errores de *cobertura* que la encuesta *telefónica*.

- e) *Precisa de más tiempo para el trabajo de campo*. Aunque en una misma fecha se envíen todos los cuestionarios a las unidades muestrales, las devoluciones de los cuestionarios rellenos no suelen ser inmediatas. La demora es habitual, como el optar por su no remisión. Además, la práctica aconsejable de enviar varios recordatorios, cada quince días (por ejemplo), para incrementar la tasa de respuesta, alarga el tiempo para la recogida de información, pudiendo superar los dos meses. Por ello no se recomienda su uso cuando se busca información inmediata y que esté afectada por la coyuntura temporal.
- f) *Mayor error de no respuesta (total y de ítem)*. Éste es su mayor hándicap, aunque no es extensible a otros modos autoadministrados: cuando la encuesta se realiza a grupos de población (estudiantes, trabajadores, clientes) reunidos en un mismo espacio. En esas circunstancias, se logra la mayor tasa de respuesta en un mínimo período de tiempo.

“Porcentajes de respuesta en el rango del 70 al 85% se ven como muy buenos, y en el rango de 60 al 70% aceptables. Porcentajes por debajo del 50% no son científicamente aceptables” (Mangione, 1998: 404-405).

Mayor *no respuesta* en personas mayores y menor nivel educativo (Kaldenberg, Koenig y Becker, 1994; Mangione, 1998; Díaz de Rada, 2002). Además, en la *encuesta por correo* es más difícil conocer el *motivo* de la *no respuesta* (rechazo, ausencia del hogar, problemas de lectura del cuestionario...) y el perfil de los no respondientes. Información que sí está disponible en las encuestas *cara a cara*. No obstante, la tasa de respuesta puede aumentarse conjugando varias de las actuaciones siguientes:

- *Elección de un marco muestral* actualizado y completo.
- *Un buen diseño del cuestionario*: no muy extenso, atractivo, fácil de comprender y rellenar por todas las personas a las que va dirigido; con espacio suficiente para responder las preguntas y con explicaciones que ayuden a su interpretación y contestación.
- *Inclusión de una carta de presentación*, que resalte la relevancia y beneficios de la participación en el estudio, además de garantizar el pleno anonimato de la información que se proporcione. También ha de incluir instrucciones precisas para la cumplimentación del cuestionario.
- *Adjuntando un sobre franqueado o de contrareembolso* para facilitar la remisión del cuestionario. Mangione (1998) recomienda sobres con sellos pegados porque los receptores no quieren “perder” sellos no devolviendo los cuestionarios. En su contra estaría su mayor coste, comparado con los sobres con contrareembolso (sólo se pagan los costes postales de los cuestionarios que son devueltos).

- *Alguna gratificación* adjuntada al cuestionario o prometiéndola tras su cumplimentación. Como se verá en el capítulo 7, las gratificaciones (económicas o materiales) previas, aunque sean de escaso valor, muestran mayor efectividad en el incremento de la tasa de respuesta que las prometidas (aunque éstas sean de mayor valor).
- *Enviando recordatorios con continuidad* a quienes no hayan devuelto el cuestionario cumplimentado (pasado un tiempo): al menos 15 días después de la entrega del cuestionario. Conviene que al recordatorio acompañe una copia del cuestionario.
- *La prenotificación* (por correo o telefónica) del envío del cuestionario puede igualmente contribuir a aumentar la tasa de respuesta. Yamarino, Skinner y Childers (1991) demuestran que su efectividad equivale a la del *recordatorio*.

Varias de estas recomendaciones se asientan en el *Método de Diseño Total* (TDM = *Total Design Method*), desarrollado por Dillman (1978, 1983, 2000) para incrementar la tasa de respuesta de las encuestas por correo. Resalta el efecto del diseño del cuestionario, del preaviso y del seguimiento de los cuestionarios (con la relevancia de que en cada envío figure el nombre de la persona a la que va dirigido).

CUADRO 5.7. Ventajas e inconvenientes de la encuesta por correo

VENTAJAS	INCONVENIENTES
<ul style="list-style-type: none"> – Abarata los costes del <i>trabajo de campo</i> – Amplía la <i>cobertura</i> de estudio: alcanza áreas aisladas y personas difíciles de localizar – El encuestado elige el momento para rellenar el cuestionario. Dispone de más tiempo para reflexionar su respuesta y comprobar la información – Ofrece mayor privacidad para responder al cuestionario. Menor vulnerabilidad al sesgo de deseabilidad social – Elimina errores de <i>medición</i> por actuación del entrevistador – Permite el uso de recursos gráficos y de preguntas con muchas opciones de respuesta 	<ul style="list-style-type: none"> – Más expuesta a errores de <i>no respuesta</i> (<i>total</i> y de <i>ítem</i>). Menor información de no respondientes – Problemas de <i>cobertura</i> en encuestas a la población general – Inasistencia al encuestado para clarificar términos complejos y estructurar la respuesta – La persona seleccionada puede no ser quien rellene el cuestionario – El encuestado puede leer todo el cuestionario antes de rellenarlo, limitando la eficacia de las preguntas de <i>control</i> y la técnica del <i>embudo</i> – Precisa más tiempo para el <i>trabajo de campo</i> – No puede recabarse información complementaria del encuestado, salvo que esté incluida en el cuestionario

Otras variedades de encuestas *autocumplimentadas* son:

- CASI (*Computer Assisted Self-Interviewing*): combina los beneficios de la entrevista personal, la encuesta autoadministrada y el uso del ordenador para la recogida de información. La persona seleccionada es quien lee y rellena el cuestionario en el ordenador que le proporciona el equipo investigador. Lo que limita su uso en personas no familiarizadas en el manejo de ordenadores. A ello se suma el coste de material informático que precisa.
- DBM (*Disk By Mail*): el cuestionario se envía en un disco por correo postal. El destinatario lo cumplimenta en su propio ordenador y lo devuelve por correo postal (en el sobre proporcionado en el envío). Su éxito depende de que el destinatario disponga de ordenador y que esté motivado para responder al cuestionario. Para evitar que el disco se deteriore con el envío postal, puede optarse por su entrega y devolución en mano (en día y hora concertados con el encuestado). Ello ayudaría a reducir la *no respuesta* (mayor dificultad de decir “no”, a la cara, a una persona; y el quedar en una fecha concreta para recoger el cuestionario genera una mayor obligación moral en el destinatario), tener información sobre motivos de la *no respuesta*, y dar instrucciones verbales para su cumplimentación.
- EMS (*Electronic Mail Survey*): la encuesta se realiza vía correo electrónico, limitando el universo de estudio a sus usuarios. El cuestionario puede proporcionarse en archivo adjunto o mediante un enlace a una página web.

Las encuestas por *correo electrónico (EMS)* y *web* presentan cuatro *ventajas* importantes:

1. *Su reducido coste* (ahorro en reprografía de cuestionarios, sobres franqueados, envíos certificados, grabación directa de las respuestas en el ordenador).
2. *Reducción del tiempo para la recogida de información* respecto al correo convencional. Además, puede conocerse, de manera inmediata, si la dirección del destinatario es correcta, si recibe el envío y lo responde. En las *encuestas web*, la obtención de respuestas es inclusive más rápida (Mann y Stewart, 2002).
3. *Mayores posibilidades gráficas y de color* (con la inclusión de gráficos interactivos).
4. Para algunas personas puede resultar *más fácil y cómodo teclear las respuestas en el ordenador* que escribirlas con bolígrafo en un cuestionario de papel.

Entre sus *inconvenientes* destacan:

- a) *Mayor dificultad de garantizar el anonimato* al encuestado.
- b) *Mayor error de cobertura*, al no ser aún universal el acceso a Internet en todos los hogares (perfil sociodemográfico del internauta aún diferencial res-

pecto del conjunto de la población: predominio de jóvenes, personas de mayor estatus social, ocupacional, de estudios y residentes en núcleos urbanos).

- c) *Errores de muestreo y no respuesta*. De acuerdo con De Leeuw y Hox (2008: 254), “la muestra en una encuesta web no es una muestra probabilística de la población en general, y no existe ningún método bueno para generar muestras aleatorias de direcciones de e-mails”.

Para la reducción del *error de no respuesta (total y de ítem)* son extensibles las recomendaciones dadas a la encuesta por *correo postal*:

- *Notificación previa del envío*.
- *Personalización del envío* (adjuntar una carta personal, con el nombre del destinatario).
- *Presentar un cuestionario atractivo*, con los mejores recursos gráficos posibles, que atraiga al destinatario y facilite su contestación.
- *Incrementar los reenvíos* (incluyendo de nuevo el cuestionario).
- *Proporcionar alguna gratificación* (aunque sea un simple “gracias”).

5.4.4. Métodos mixtos de encuesta

Cada vez son más las voces que abogan por la conjunción de métodos de encuesta en un mismo proyecto de investigación. Llega incluso a afirmarse que “los sistemas de modos mezclados son la norma hoy en día” (Biemer y Lyberg, 2003: 208-209). Si bien, ya la defendió Stanley Payne en 1964 (“Combination of survey methods”, *Journal of Marketing Research*, 1: 61-62). A partir de los años ochenta comienza su expansión, adoptándose importantes encuestas gubernamentales en diferentes países (*U.S. Current Population Survey*, *British Labour Force Survey*, *Danish Omnibus and Labour Force Survey* o *Finnish Labour Force Survey*, por ejemplo). Varias de ellas son *encuestas panel*, que comienzan con una *entrevista personal* (primer contacto) y *telefónica* o *autocumplimentada* para los siguientes, reduciendo los costes económicos de la investigación.

Pero no es sólo la finalidad económica la que propicia el uso de métodos mixtos. También la reducción de errores de *no respuesta*, de *cobertura* y de *medición*, como lo ilustran las siguientes combinaciones:

- *Encuesta personal para completar la información no recabada por las encuestas telefónicas o autoadministradas*. Permite reducir el error de *no respuesta*, siguiendo a los no respondientes de los otros métodos, además de posibilitar la obtención de información sobre el perfil de los “no respondientes” (imprescindible para valorar los efectos de la *no respuesta* en la calidad de la encuesta). Igualmente contribuye a reducir el error de *no cobertura* (personas sin acceso telefónico, Internet...).

- *Encuesta telefónica para reducir la no respuesta en los otros métodos.* Llamar por teléfono a las unidades de la *muestra* antes de la encuesta *cara a cara* o autoadministrada, con el propósito de facilitar el *trabajo de campo* (motivarles y resaltar la relevancia de su contribución para la investigación). También puede aplicarse después, en el seguimiento de los “no contactados” por los otros métodos de encuesta (en los “no contactados” *cara a cara*, concertar una “cita” para la realización de la entrevista personal; en caso de que no puedan, cabe la opción de su realización por teléfono).
- *Encuesta autocumplimentada para reducir la no respuesta en los otros métodos.* Primer contacto vía *correo* (postal, electrónico) solicitando su cooperación e indicando la fecha de la *entrevista personal* o *telefónica*. Después, enviando por *correo* el cuestionario a las personas no contactadas *personal* o *telefónicamente*.
- *Encuesta cara a cara con hojas de respuesta autocumplimentada* para reducir los errores de *medición* debidos al sesgo de la deseabilidad social (como ya demostraran Sudman y Bradburn, 1974 y 1982/1987). La autocumplimentación de *hojas de respuesta* es habitual que se haga “en presencia” del entrevistador, salvo que sea una parte considerable del cuestionario. En este último caso, puede concertarse día y hora para la recogida posterior del cuestionario, con preferencia al envío postal en sobre franqueado (porque puede no devolverlo o en plazo). La *autoadministración* de parte del cuestionario también es extensible a la encuesta *telefónica*. En este caso, al encuestado se le comunica que va a recibir el cuestionario por correo (o en mano) para su autocumplimentación. Además de favorecer la respuesta en temas sensibles, permite el uso de *ayudas visuales*. La conjunción de entrevista con autocumplimentación también se realiza en las encuestas *diario*.

La combinación de diferentes modos de encuesta persigue la consecución de dos beneficios principales:

- a) *Reducir errores de encuesta:* combinado diferentes *marcos muestrales* (error de *cobertura*), procedimientos de selección de las unidades de la *muestra* (error de *muestreo*) y de administración de los cuestionarios (errores de *medición* y de *no respuesta*). Como señalara Lavrakas (1998: 449), “El error total de la encuesta puede reducirse si las limitaciones de un modo son contrarrestadas por las fuerzas del otro”.
- b) *Disminuir costes* (tiempo y dinero): aplicar métodos menos caros (encuesta por *correo*, *telefónica*), tras la encuesta *cara a cara*, para el seguimiento de los *no respondientes*.

Entre los *inconvenientes* destacan:

- a) No todos los centros de opinión pública disponen de *medios y personal preparado para la aplicación combinada de la encuesta personal y telefónica*.

- b) *Problemas en el diseño de la muestra*: incompatibilidades de los *marcos muestrales* (diferente grado de *cobertura*), y los debidos a la conjunción de diferentes procedimientos para la selección de la *muestra*.
- c) *Problemas en el diseño del cuestionario*, salvo que el cuestionario sea sencillo y pueda aplicarse sin distinción por método de encuesta (en el capítulo 6 se abordan las peculiaridades del cuestionario en cada método de encuesta).
- d) *Problemas de comparabilidad de las respuestas*, sobre todo en temas subjetivos y vulnerables al sesgo de la deseabilidad social. Cómo se aplique el cuestionario afecta a la respuesta.
- e) *Problemas de coste añadido* (en tiempo y dinero), cuando se opta por la aplicación de otro método para la captación de *no respondientes* (comparado con la opción la alternativa de la *sustitución* del individuo “no localizado” por otro similar).

Lecturas complementarias

- Alvira, F. (2004/2011): *La encuesta: una perspectiva general metodológica*. CIS. Cuaderno Metodológico nº 35. Madrid.
- Battaglia, M. P.; Link, M. W.; Frankel, M. R.; Osborn, L. y Mokdad, A. H. (2008): “An evaluation of respondent selection methods for household mail surveys”. *Public Opinion Quarterly*, 72 (3): 459-469.
- De Leeuw, E. D.; Hox, J. J. y Dillman, D. A. (comps.), (2008): *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York.
- Groves, R. M.; Fowler, F. J. Jr.; Couper, M. P.; Lepkowski, J. M.; Singer, E. y Tourangeau, R. (2009): *Survey methodology*. Wiley & Sons. Hoboken, New Jersey.
- Kreuter, F.; Presser, S. y Tourangeau, R. (2008): “Social desirability bias in Cati, IVR, and Web Surveys. The effects of mode and question sensitivity”. *Public Opinion Quarterly*, 72 (5): 847-865.
- Lepkowski, J. M.; Tucker, C.; Brick, J. M.; De Leeuw, E. D.; Japec, L.; Lavrakas, P. J.; Link, M. W. y Sangster, R. L., (2008): *Advances in telephone survey methodology*. Willey. Nueva York.

Ejercicios propuestos

1. Destaque tres aportaciones pioneras en el desarrollo de la encuesta. Justifique su elección.
2. Localice una investigación donde se articule la encuesta con otras estrategias de investigación. Describa la función que cada una cumple en la investigación, sus posibilidades y sus límites.
3. Concrete los *errores* a los que está más expuesto cada método de encuesta y las actuaciones para su reducción.
4. Destaque las contribuciones de los nuevos avances tecnológicos para la mejora de los métodos de encuesta y cómo afectan a la reducción de cada error de encuesta.
5. Elija un problema de investigación y seleccione el método o la combinación de métodos de encuesta que aplicaría para su resolución. Justifique la respuesta y diseñe la investigación.

6

EL DISEÑO Y LA EVALUACIÓN DE CUESTIONARIOS

El cuestionario es el instrumento básico de recogida de información en la investigación mediante *encuesta*. Aunque también está presente en otras estrategias de investigación, como la *experimentación* o la *investigación no reactiva* (para el registro de datos observados o incluidos en informes o expedientes). En el capítulo 5 ya se señaló que de su diseño u aplicación depende, en gran medida, la calidad de la encuesta. A lo largo del presente capítulo se van a ir desgranando los diferentes *errores* de medición debidos al diseño del cuestionario. A los errores siguen actuaciones o recomendaciones encaminadas a su reducción. El recorrido concluye con la panorámica de procedimientos de evaluación del cuestionario. Todo ello encaminado a la mejora de su diseño y aplicación.

6.1. Elaboración del cuestionario

La obtención de datos que se correspondan a la “realidad” es el reto al que se enfrenta cualquier investigación. Cuando los datos provienen de las respuestas (orales o escritas) recabadas de personas, el reto se amplía. Que la respuesta no se ajuste a la realidad puede deberse a la voluntad de la persona de no proporcionar una respuesta veraz. Las personas no siempre decimos la verdad. En especial en cuestiones que afectan a nuestra imagen personal y pública, al *qué dirán*. Pero que se proporcionen respuestas erróneas puede ser también involuntario. Puede deberse a una mala redacción de la pregunta, que dificulte su comprensión. También, a desconocimiento del tema sobre el que se pregunta o a problemas de recuerdo o memoria. Errores a cuya aminoración puede contribuir un buen cuestionario.

El *cuestionario* consiste en un listado de preguntas estandarizadas, cuya secuencia y formato de pregunta-respuesta no varía para cada unidad que compone la

muestra de estudio. Ello posibilita la comparación de las respuestas: que éstas se deban a diferencias de los respondientes, y no a divergencias en la formulación de las preguntas, las opciones de respuesta o su ubicación en el cuestionario. La figura 6.1 sintetiza errores debidos a la formulación de la pregunta. Éstos se dividen en:

- a) *Errores de especificación*. La pregunta no se adecua a los objetivos de la investigación, no habiendo correspondencia con el concepto que trata de medir. No logra su traducción en un indicador válido para su medición. Éste es el primer criterio que actúa en la evaluación de un cuestionario. De ahí que sea prioritario, antes de comenzar a redactar preguntas, decidir qué conceptos son claves para medir el problema de estudio y cómo van a operacionalizarse.
- b) *Errores de medición* debidos a la *desigual interpretación* de términos de la pregunta, por el empleo de palabras ambiguas o con diferentes connotaciones. De especial incidencia en cuestionarios autocumplimentados, que no se benefician de la presencia de un entrevistador (que pueda aclarar el significado de la pregunta y la respuesta). Como indicase Groves (1989: 450), “Aunque el lenguaje en las preguntas de encuesta puede estar estandarizado, no existen garantías de que el significado asignado a las preguntas sea constante por encima de los encuestados”. Por lo que ha de evitarse que tanto la pregunta como las opciones de respuesta (en preguntas *cerradas*), incluyan palabras ambiguas o que puedan tener una connotación dispar.
- c) *Errores de medición* por la *redacción* de la pregunta, su introducción y formulación exacta. Errores “invisibles al analista” (Groves, 1989: 497), a menos que disponga de información al respecto.
- d) *Errores de medición* por las *opciones de respuesta*. Afecta qué alternativas de respuesta se ofrezcan (en preguntas *cerradas*), si se presentan de manera equilibrada (que no induzca a una respuesta concreta), y el *orden* en que se presentan. Sobre todo cuando son varias (más de tres) las alternativas de respuesta a elegir y su disposición propicia la elección preferente de una de las opciones de respuesta: la primera (efecto de *primacía*) o la última (efecto de *recencia*), con independencia del contenido de la respuesta. Ambos efectos se dan más en personas con un bajo nivel de estudios (Schuman y Presser, 1981/1996; Krosnick y Alwin, 1987; Krosnick y Presser, 2010), al igual que el sesgo de *acquiescencia* (o tendencia al acuerdo indistintamente de lo que se pregunte)
- e) *Errores de medición* debidos al *orden* y el *contexto* de la pregunta en el cuestionario. Se deben a la transferencia de significados de una pregunta a otra. Ya en 1939 J. Sayre comprobó (“A comparison of three indexes of attitudes towards radio advertising”. *Journal of Applied Psychology*, 23: 23-33) que las preguntas precedentes pueden influir en las respuestas a preguntas posteriores. Varios de sus hallazgos fueron con posterioridad corroborados en experimentos realizados por Cantril (1944) o Schuman y Presser (1981/1996). Las

preguntas generales (más que las específicas) son las más vulnerables a este error, como después se verá.

- f) *Errores de medición* atribuidos al *tema* de la pregunta, que inhiban de dar una respuesta veraz: preguntas “embarazosas” (Goode y Hatt, 1952), “amenazantes” (Sudman y Bradburn, 1974), “sensitivas” (Lin, 1976) u “ofensivas” (Berdie y Anderson, 1974), que incomoden al encuestado. Pero también porque se hallen referidas a acontecimientos de “pasado”.

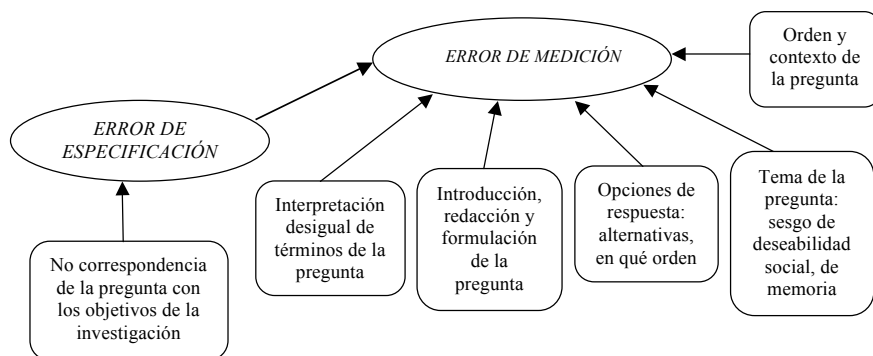


FIGURA 6.1. Errores debidos a la formulación de la pregunta.

Cualquiera de estos errores puede traducirse en error *sistemático* y *aleatorio*, dependiendo de su incidencia en el conjunto de la encuesta.

- *Errores sistemáticos*: aquellos que “no suman cero cuando las observaciones de la muestra son promediadas” (Biemer y Lyberg, 2003: 47). Resultan en una sobreestimación o subestimación de los valores “reales” en la población. Por ejemplo, los sesgos de *acquiescencia* y de *deseabilidad social*.
- *Errores aleatorios*: afectan menos a la calidad de la encuesta, al no generar una sobreestimación o subestimación de los valores promedio. Caso de la pregunta acerca del nivel de ingresos, que puede suscitar una declaración que excede el nivel “real” (para “impresionar”) o inferior, pero no una decantación preferente por uno u otro. Por lo que su efecto puede reducirse incrementando el tamaño de la *muestra*. No así el correspondiente a los *errores sistemáticos*, que precisan de un mayor cuidado cuando se diseña la encuesta y el cuestionario.

6.1.1. Actuaciones previas a la redacción de preguntas

Varios de los errores de *medición* (o respuesta) pueden reducirse con un buen diseño de cuestionario. Pero, como se señalara en el capítulo 5, antes de comenzar a redactar preguntas conviene cumplir una serie de actuaciones previas:

- a) *Indagación exploratoria cualitativa* (observación, grupos de discusión, entrevistas abiertas) para descubrir la variedad de discursos y aspectos que incluye el problema de investigación. De mayor relevancia cuando se quiere generar “nuevas” preguntas.
- b) *Revisión bibliográfica* de aspectos teóricos y empíricos relativos al problema de investigación. Acceder a cuestionarios que ya han sido “validados” (en estudios anteriores) adquiere una especial relevancia. Ver qué preguntas se hicieron, qué trataban de medir, qué resultados obtuvieron, puede ayudar a la redacción de nuevas preguntas, evitando caer en errores cometidos con anterioridad.

El investigador puede también optar por tomar algunas preguntas, mediante las que se alcanzó una buena medición de los conceptos de interés, e incluirlas en su cuestionario. De esta forma pueden realizarse análisis comparativos y de tendencias, además de constatar su *fiabilidad*. Para ello no han de modificarse las preguntas tomadas de otros cuestionarios (su redacción y ubicación en el cuestionario), y siempre citando la fuente original.

Si la pregunta procede de cuestionarios redactados en otro idioma o en el mismo, pero en distinto contexto sociotemporal, habrá que revisar su redacción: comprobar si incluye términos que en el actual contexto tendrían una connotación distinta.

- c) *Entender los procesos sociopsicológicos* implicados en la respuesta a preguntas de encuesta ayuda a la formulación de “buenas” preguntas. Por lo que cada vez son más los metodólogos que asientan el diseño de cuestionarios en teorías cognitivas y de la comunicación personal.

A las recomendaciones anteriores se añaden las siguientes como actuaciones previas a la redacción de preguntas:

- a) *Especificar los objetivos* de la encuesta, la información que pretende recogerse. Cada pregunta ha de adecuarse a los objetivos del estudio. Ante cada pregunta nos deberíamos preguntar: “¿por qué hago esta pregunta?” (Sudman y Bradburn, 1987). En caso de no ajustarse, conviene eliminarla. Para ello son de gran utilidad las “hojas de resumen”. Adviértase que uno de los criterios de evaluación de las preguntas es su adecuación a los objetivos de la encuesta.
- b) *Identificar los conceptos* centrales en la investigación y traducirlos en *indicadores* “válidos”, que ayuden a reducir los errores de *especificación*. Se concreta en lo señalado en el capítulo 2:
 - *Seleccionar qué conceptos son claves* para medir el *problema* de investigación. Conviene diferenciar lo que el concepto “incluye” y “excluye”; las diferencias con otros conceptos relacionados; y describir su opuesto (ayuda a comprender lo que incluye y excluye). *Por ejemplo*: tolerancia-intoleran-

cia, xenofobia-xenofilia, discriminación-integración, satisfacción-insatisfacción.

- *Descomponer el concepto en dimensiones* (o aspectos que incluye). Preferible acotar las más relevantes que abarcar indiscriminadamente muchas dimensiones.
 - *Buscar indicadores* (para cada dimensión) que logren representar la propiedad latente del concepto que miden. Es preferible acotar el número de dimensiones y aumentar el de indicadores que a la inversa. Aumenta la validez de constructo, si se alcanzan los mismos resultados con indicadores diferentes.
- c) *Elegir el método de encuesta* y concretar la *población* a sondear. Ha de tenerse presente el nivel de formación de la población, su dominio del idioma, su capacidad de comunicación verbal, además de su interés por el tema. Como regla se impone el utilizar palabras que sean comprendidas por aquellos de menor nivel de formación en la población de estudio. Se trata de elaborar un cuestionario que le sea atractivo, fácil de responder y que no lleve a su rechazo. Asimismo, el modo en que se administrará el cuestionario influye en su formato, presentación y la redacción específica de las preguntas, como después se verá.
- d) *Concretar los análisis* que se prevé realizar, en consonancia con los objetivos de la investigación. Como se verá en el capítulo 8, cada técnica analítica tiene unos supuestos básicos que garantizan su correcta realización. En la medida en que las variables incluidas en el cuestionario los satisfagan, los análisis podrán llevarse a buen término. Por ello se recomienda incluir variables “relevantes” para el objeto de estudio y que se hallen medidas en el nivel correcto que amplíe sus posibilidades analíticas.

Por ejemplo, la variable *edad* en años cumplidos (con preferencia a su medición ordinal e incluso en intervalos de edad) o la *ideología política* en una escala de diez puntos en vez de tres o cinco. Sirva de ilustración cómo están ambas variables medidas en las encuestas del CIS (Centro de Investigaciones Sociológicas):

- “¿Cuántos años cumplió Ud. en su último cumpleaños?” _____
- “Cuando se habla de política se utilizan normalmente las expresiones izquierda y derecha. En esta tarjeta hay una serie de casillas que van de izquierda a derecha. ¿En qué casilla se colocaría Ud.?”

Izqda.					Dcha.					
00	01	02	03	04	05	06	07	08	09	10

6.1.2. La redacción de las preguntas

¿Qué es una “buena” pregunta? “Aquella que no afecta a la respuesta” (Payne, 1980: 72). “Aquella que produce respuestas que son medidas válidas y fiables de algo que queremos medir” (Fowler, 1998: 344): *válidas* en el sentido de que se correspondan con el valor “verdadero” que intentan medir; *fiables* en la medida en que las respuestas sean consistentes (capítulo 2). Aunque algunos consideren la elaboración de preguntas un “arte” que se adquiere por experiencia (el mismo Stantley Payne publica en 1951 un pequeño libro con el título *The Art of Asking Questions*), cada vez son más los metodólogos que asientan el diseño de cuestionarios en teorías cognitivas y de comunicación interpersonal: entender los procesos socio-psicológicos implicados en la respuesta ayuda a la formulación de “buenas” preguntas.

“Qué difícil es escribir buenas preguntas”. “Requiere gran habilidad y juicio, mucho tiempo y experiencia” y, todavía, “incluso después de años de experiencia, nadie puede escribir una pregunta perfecta” (Sudman y Bradburn, 1982: 3, 283).

El cuadro 6.1 resume las *recomendaciones* más citadas en la literatura de encuesta para la *redacción* de preguntas. Cada una de ellas se explicita a continuación, si bien agrupadas de acuerdo con el *error de medición* que muestran reducir.

CUADRO 6.1. Recomendaciones para la redacción de preguntas

-
- Formular preguntas relevantes para la investigación
 - Preguntas breves y fáciles de comprender por las personas a las que van dirigidas
 - Emplear un vocabulario sencillo, evitando conceptos o términos abstractos
 - La redacción de la pregunta ha de ser precisa
 - La pregunta ha de ser lo más objetiva o neutra posible para no influir en la respuesta
 - No redactar preguntas en forma negativa
 - Redactar preguntas de forma personal y directa
 - Ni las preguntas ni las alternativas de respuesta (cuando se ofrezcan) pueden referirse a varias cuestiones al mismo tiempo. El principio de la *idea única*
 - No emplear palabras que comportan una reacción estereotipada
-

Recomendaciones para reducir errores de medición debidos a la interpretación de términos de la pregunta (por el empleo de palabras ambiguas o con diferentes connotaciones):

- a) *Emplear un vocabulario sencillo, evitando conceptos o términos abstractos*, sobre todo cuando se encuesta a personas mayores o con problemas de lenguaje. Las preguntas complejas o difíciles de entender suelen obtener un ma-

porcentaje de *no respuesta de ítem* (no saben/no contestan). La elección de las palabras ha de estar supeditada a las peculiaridades de la población. Como regla, utilizar palabras que sean comprendidas por los miembros de menor nivel educativo.

Términos aparentemente sencillos, como “proporción”, pueden provocar interpretaciones erróneas, como lo muestra Belson (1981: 240). A la pregunta “¿Qué proporción del tiempo que ve la televisión lo dedica a ver programas de noticias?” sólo 14 de las 53 personas entrevistadas interpretaron correctamente la palabra “proporción” como porcentaje, parte o fracción. El resto lo interpretó como “cuánto tiempo” o “qué programas de noticias ve”.

“Cuando los respondientes no entienden las palabras utilizadas en las preguntas, podrían pedir clarificación, podrían negarse a contestarla, o podrían adivinar el significado de la frase desconocida y responder la pregunta. Todas estas tres opciones disminuyen la fiabilidad de la medición” (Fowler y Consenza, 2008: 139).

- b) *Incluir palabras familiares*, que se empleen con frecuencia, pero siempre que no tengan demasiados significados o diferentes connotaciones. *Por ejemplo*, “dolor de cabeza” en lugar de “cefaleas”. Especial relevancia adquiere en cuestionarios autocumplimentados sin la presencia de un entrevistador, que pueda aclarar el significado de la pregunta y las respuestas. Se recomienda, ante la duda, consultar en el diccionario si la palabra elegida tiene otros significados que puedan confundir al encuestado sobre el sentido de la pregunta. También hay que ser precavido cuando se emplean términos como trabajador, mayor, joven, progresista, mucho, con bastante frecuencia, usualmente, barato, pues su acepción puede variar dependiendo de a quién se pregunte y dificulta la interpretación de la respuesta.
- c) Para evitar su distinta interpretación Fowler (1998) recomienda no utilizar *escalas con adjetivos* para medir estados subjetivos (actitudes, opiniones, valores), en estudios internacionales o que incluyan a poblaciones de culturas diversas. En estos casos se propone el *empleo de escalas numéricas* para garantizar la comparabilidad de las respuestas.
- d) En encuestas a poblaciones de diferente procedencia cultural o idioma, conviene asegurarse de que la traducción a otro idioma no cambie el significado de la pregunta. En caso contrario, no podrían compararse las respuestas, como sucedió en algunas de las preguntas incluidas en el *Eurobarómetro 53*, del año 2000, que presentaron *sesgos culturales* (diferencias de interpretación), como se viera en el capítulo 5.

Una vez redactada la pregunta, conviene ponerse en el lugar de la persona de menor nivel educativo a encuestar y preguntarse: “¿Cómo respondería yo a la pregunta?”. Después, en el *pretest* o precomprobación del cuestionario, el investigador tendrá la oportunidad de comprobar si realmente la pregunta es entendida por la

población a la que se dirige. Se fijará en los problemas habidos en su administración y la proporción de personas que no responden a la pregunta (los “no saben/no contestan”). Ello exige que el *pretest* se realice con personas que se ajusten a toda la variedad de unidades que forman la población de estudio (apartado 6.2).

Recomendaciones para reducir errores de medición debidos a la redacción de la pregunta, su introducción y formulación concreta:

- a) *Preguntas breves y fáciles de comprender*, por las personas a las que van dirigidas. De acuerdo con Converse y Presser (1994: 93), “la mejor estrategia es utilizar preguntas breves, cuando sea posible, y que los entrevistadores den a los entrevistados tiempo suficiente para responder a las preguntas”. Las *preguntas largas y complejas* aumentan la probabilidad de que el encuestado se pierda y no las siga (Orenstein y Phillips, 1978; Molenaar, 1982).

Por ejemplo, sería desaconsejable emplear preguntas como: “*Algunas personas opinan que el Gobierno no debería ceder a las presiones nacionalistas, otros dicen que es el momento de reformar la Constitución, ¿qué piensa usted?*”

Excepto cuando quiera reducirse el sesgo de la “deseabilidad social” y mejorar el recuerdo (en preguntas referidas al pasado), como se verá después.

Así se constató en experimentos realizados por Bradburn, Sudman y colaboradores (1979), con preguntas como, *por ejemplo*, “*Ocasionalmente, la gente bebe con el estómago vacío o bebe demasiado y se maree. El año pasado, ¿con qué frecuencia se mareó mientras bebía algún tipo de bebida alcohólica?*”. El preámbulo contribuyó a una mayor declaración de consumo de bebidas alcohólicas. Al igual que sucedió en preguntas sobre conducta sexual.

De acuerdo con Sudman y Bradburn (1982) las preguntas *largas* se prefieren para el análisis de conductas (no de actitudes), y pueden utilizarse para estimular el recuerdo. Llevan más tiempo para que el entrevistador las lea, proporcionando al entrevistado más tiempo para pensar y meditar su respuesta. Además, la longitud de la pregunta puede tomarse como un indicador de su importancia. Anima a prestar mayor atención a la pregunta (Sudman y Bradburn, 1974, 1983; Cannel *et al.*, 1981; Kalton y Schuman, 1982/2002).

Por ejemplo, Cannel, Miller y Oksenberg (1981) comparan dos formulaciones (corta y larga) de una misma pregunta:

- Corta: “¿Qué problemas de salud ha tenido usted en el pasado?”.
- Larga: “La siguiente pregunta es sobre problemas de salud durante el último año. Esto es algo que preguntamos a todo el mundo en la encuesta. ¿Qué problemas de salud ha tenido usted el año pasado?”

Obtuvieron un mayor relato de problemas de salud con la pregunta *larga*.

Por lo que, como puntualizan Kalton y Schuman (1982/2002: 15), “el consejo usual de “mantén las preguntas cortas” probablemente es una forma inadecuada de decir “mantén las preguntas sencillas”; en la práctica, las dificultades de las preguntas largas probablemente derivan de su complejidad más que de su longitud en sí misma”.

- b) *La redacción de la pregunta ha de ser precisa.* Cuantos más detalles aporte, más facilita su comprensión e interpretación.

Hechos aparentemente sencillos como, *por ejemplo*, “*número de habitaciones ocupadas en una vivienda*”, puede tener problemas de definición: ¿Incluye la sala de estar, la cocina, los baños, el hall...? ¿Una habitación dividida por un biombo es una o son dos habitaciones?... Para evitar una interpretación errónea de la pregunta conviene añadir alguna especificación que concrete el sentido de la pregunta. Otros *ejemplos*:

- “Aproximadamente, *¿cuánto dinero se gastó usted en las pasadas navidades?*”. (Inclúyase el dinero gastado *en todo tipo* de conceptos: alimentación, transporte, viajes, compras de regalos, lotería y demás).
- “*¿Cuántos años cumplidos tiene Ud.?*”
- “Actualmente, entre todos los miembros del hogar (incluido el entrevistado) y por todos los conceptos, *¿de cuántos ingresos netos disponen por término medio en su hogar al mes?*”

- c) *La pregunta ha de ser lo más objetiva o neutra posible para no influir en la respuesta.* Aunque autores, como Hox (1997: 53), afirmen que “no existe ninguna pregunta de encuesta puramente objetiva”, la búsqueda de la “objetividad” se convierte en meta, al ser un requisito de calidad de la pregunta. Como dice Payne (1980: 72), “Una buena pregunta es, entre otras cosas, aquella que no afecta a la respuesta”. Han de evitarse preguntas tendenciosas que inciten a una respuesta concreta. También, preguntas *cerradas*, que fuercen a los encuestados a elegir entre pocas opciones de respuesta (no cubriendo todo el abanico de alternativas posibles). “Especificar sólo una alternativa de respuesta es una forma de cargar la pregunta” (Sheatsley, 1983: 214). *Por ejemplo*:

- “*¿No pegará usted a sus hijos, verdad?*”. Incitaría una respuesta negativa y no aportaría una información creíble.
- “Recientemente se ha propuesto aumentar el gasto sanitario para dar asistencia médica a todos los colectivos de inmigrantes. *¿Le parece a usted una buena propuesta?*”. Es probable que capte un estado de opinión más favorable a la propuesta que cuando el enunciado de la pregunta incluya: “*¿Le parece a usted ésta una buena o una mala propuesta?*”
- “*¿Está usted a favor o en contra de la pena de muerte para personas condenadas por terrorismo?*” mejor que sólo preguntar: “*¿Está Ud. a favor de...?*”.
- Lo mismo puede decirse de la pregunta “*¿Cree usted que la actual política exterior del Gobierno es positiva?*”. Suscitara más respuestas favorables a la

actuación gubernamental que cuando el enunciado de la pregunta incluye los términos: “*excelente, buena, regular, deficiente, mala*”.

La pregunta ha de ser equilibrada en consideraciones positivas y negativas, si quiere evitarse el *sesgo de aquiescencia*. Sobre todo conociendo la predisposición de algunas personas a manifestarse “de acuerdo” con cualquier proposición plausible que se le ofrezca.

También conviene evitar expresiones como “*la mayoría de la gente opina que...*”; o la referencia expresa a una autoridad (la Iglesia, el Gobierno, los expertos) precediendo a la pregunta, salvo que deliberadamente se elija como estrategia para aumentar la probabilidad de informar sobre conductas ilícitas o socialmente censurables. Como después se verá, estas introducciones tienen el efecto de “cargar” la pregunta, influyendo en la respuesta.

“Muchos expertos creen que la actuación del Gobierno en el control de la inmigración no ha sido adecuada, pero otras personas opinan que ha sido acertada. ¿*Qué opina Ud.*?”

Para evitar sesgos en la pregunta conviene discutir su redacción con personas que compartan otros puntos de vista (que actúen de “abogados del diablo”). Ello contribuye a enriquecer la pregunta.

- d) *No redactar la pregunta con una negación* porque dificulta su comprensión. De manera especial, cuando se pide al encuestado que manifieste su grado de acuerdo o desacuerdo con determinadas cuestiones. En caso que se prefiera la redacción negativa de la pregunta, conviene destacar la palabra subrayando “**No**” para llamar la atención y que la persona se percate del sentido de la pregunta.

Por ejemplo, en vez de la pregunta “¿*No debería castigarse a los estudiantes que suspenden?*”, es preferible: “*A los estudiantes que suspenden, ¿se les debería castigar?*”. Este último enunciado clarifica el significado de la respuesta negativa.

Asimismo evitar ítems como los incluidos en encuestas sobre familia y género, realizadas por el CIS dentro del *International Social Survey Program*: “*No está bien que la madre se vaya a trabajar y el padre se quede en casa cuidando de los hijos*” o “*Un matrimonio no debería separarse aunque no se lleve bien, incluso si no tiene hijos*”. En ambos casos se pide el grado de acuerdo (Muy de acuerdo, De acuerdo, En desacuerdo, Muy en desacuerdo). Lo que hace más exigible su formulación en afirmativa para facilitar la comprensión e interpretación unívoca de la pregunta.

- e) *Redactar la pregunta de forma personal y directa*. Las preguntas específicas suelen proporcionar una información más precisa que las genéricas (más vulnerables a mayor variedad de interpretaciones). Razón por la cual se aconseja (Converse y Presser, 1994) restringir su uso a aquellas circunstancias donde de interés obtener una medida global.

Por ejemplo, Molenaar (1982) cambió la pregunta “¿Está usted de acuerdo con...?” por una pregunta más directa, “¿Votaría usted por...?”, obteniendo de un 4 a un 12% más de acuerdo con la política gubernamental.

O cambiar la pregunta genérica “¿Alguien le atacó a usted en los últimos seis meses?” por una serie de preguntas específicas que detallen diferentes maneras de ser atacado (con pistola, cuchillo, bate de béisbol, satén, palo...), como se ha hecho en *The National Crime Survey* de Estados Unidos.

En la *Encuesta de Actitudes ante la Inmigración* realizada por el CIS en 1996 (estudio n.º 2.214), comenzaba por la pregunta “Para empezar, ¿cree Ud. que, en general, toda persona debería tener libertad para vivir y trabajar en cualquier país, aunque no fuera el suyo?”. El 95% de los 2.491 encuestados respondió “sí”. Porcentaje que contrasta con el 64% que después se posicionó muy o bastante de acuerdo con el establecimiento de un cupo anual para la entrada de inmigrantes (“El Gobierno español ha establecido un cupo anual para la entrada de inmigrantes, esto es, un número máximo de extranjeros que pueden entrar a trabajar en el país cada año. En principio, ¿está usted muy, bastante, poco o nada de acuerdo con esta medida?”). A efectos analíticos es de mayor interés esta pregunta que la anterior.

Asimismo, si se analizan conductas que precisan un detalle temporal o numérico, conviene proporcionar categorías de respuesta específicas.

Expresiones como “con frecuencia” o “regularmente” son vagas y ambiguas. Habría que evitarlas. Son preferibles opciones de respuesta concretas.

Por ejemplo, “diariamente”, “2-3 veces por semana”, “una vez a la semana”, “dos veces al mes”.

f) *Ni las preguntas ni las alternativas de respuesta pueden referirse a varias cuestiones al mismo tiempo.* Se defiende el principio de idea única.

Por ejemplo, preguntas como “¿Cree usted que la sociedad actual es egoísta e insolidaria?” o “¿Le gustaría ser rico y famoso?”, no son adecuadas. No ayudan a una interpretación correcta de la respuesta, al no poderse separar la doble mención de la pregunta.

Lo mismo es extensible a la pregunta incluida en las encuestas CIRES/ASEP de actitudes ante la inmigración: los encuestados han de elegir la afirmación con la que se sientan más de acuerdo:

- “Que los inmigrantes vengan a España con sus familiares y se asienten para siempre”.
- “Que los inmigrantes vengan una temporada sin sus familiares y que luego regresen a su país”.

Así expresadas, ambas opciones de respuesta incluyen dos cuestiones a la vez. Una, el carácter familiar e individual de la inmigración. Dos, su carácter definitivo o temporal. De esta manera se dificulta la interpretación de las

respuestas. Además se fuerza a los encuestados a elegir entre dos opciones de las cuales pueden compartir una parte, pero no la otra. Mejor sería haber ampliado las opciones de respuesta para que cada una de ellas incluya una única idea de la inmigración que se prefiere.

Otro *ejemplo* lo proporciona la pregunta incluida en las encuestas telefónicas de actitudes ante la inmigración de IESA-OBERAXE (Pérez Yruela y Desrues, 2007): “¿Cree Ud. que es más bien positivo o negativo para la sociedad que esté compuesta por personas de origen racial, religión y cultura diferentes?”. En la encuesta *cara-cara* CIS-OBERAXE de 2008 y siguientes fue modificada por preguntas específicas originales de la encuesta experimental *MEXEES I* (Cea D’Ancona y Valles, 2010a):

“En una escala de 0 a 10, donde 0 significa “muy negativo” y 10 “muy positivo”, valore la conveniencia de que la sociedad española esté compuesta por personas de diferentes... (MOSTRAR TARJETA)

	Escala		
	00 - 10	NS	NC
– Países	_____	98	99
– Culturas	_____	98	99
– Religiones	_____	98	99
– Razas	_____	98	99

También se constata (Cea D’Ancona y Valles, 2010b) que el empleo del término “persona” propicia una respuesta más favorable que el de “inmigrante”.

6.1.3. El formato de la pregunta

La pregunta puede figurar en el cuestionario con un formato *abierto* o *cerrado*, dependiendo de si aparecen expresas las alternativas de respuesta.

Preguntas abiertas: aquellas que no especifican las opciones de respuesta. El encuestado puede expresarse con sus propias palabras, no circunscribiendo su respuesta a una alternativa predeterminada.

EJEMPLO DE PREGUNTAS ABIERTAS

¿Cuál es, a su juicio, el principal problema que existe actualmente en España? _____

Cuando oye la palabra inmigración, ¿qué es lo primero que le viene a la mente? _____

La inmigración, como todo, tiene aspectos positivos y negativos. ¿Qué aspecto positivo destacaría más? _____

¿Cuál es/era su actual/última ocupación u oficio? Es decir, ¿en qué consiste/tía específicamente su trabajo? (Precisar lo más posible las actividades realizadas. EJEMPLO: mecánico reparador de automóviles, ayudante de odontología, profesor de enseñanza primaria, etc.) _____

Esta “libertad” de respuesta conlleva unas ventajas y unos inconvenientes. Entre las *ventajas* principales están:

- a) *Más fáciles de elaborar que las preguntas cerradas*, al no exigir mucho conocimiento sobre lo que se pregunta.
- b) *Ocupan menos espacio* en el cuestionario.
- c) *Permiten a los encuestados responder con sus propias palabras*, no sugiriéndoles ningún tipo de respuesta. Por el contrario, cuando se les pide que respondan eligiendo una respuesta predeterminada, pueden sentirse forzados a seleccionar una respuesta que no se ajusta a la que darían si la pregunta fuese *abierta*. Además, pueden introducir “matices”, que ayuden a interpretar su respuesta.
- d) *Pueden obtenerse respuestas no previstas o inesperadas*, cuando se diseñó el cuestionario.
- e) *Permiten captar lo más “saliente” para el encuestado*. Lo primero que le viene a la mente.

Preguntas como, *por ejemplo*, “¿Qué piensa sobre...?”, “¿Por qué?”, conceden plena libertad al encuestado para pronunciar cualquier idea que le surja. Si el cuestionario es rellenado por un entrevistador, hay que insistirle que escriba exactamente lo que dice el encuestado; que no introduzca ninguna modificación a su respuesta.

Pese a sus amplias posibilidades, las preguntas *abiertas* no predominan en los cuestionarios. Se debe a la confluencia de los siguientes *inconvenientes*:

- a) *Más caras que las preguntas cerradas*. Requieren un mayor esfuerzo en su registro (por parte del entrevistador y el encuestado), transcripción y codificación. Como indica Payne (1980: 53-54): “Pocas personas emplean las mismas palabras incluso para expresar la misma idea. Algunos clarifican sus afirmaciones mejor que otros. Esto hace la grabación difícil cuando queremos distinguir entre dos ideas estrechamente relacionadas pero aun así diferentes”. Esta dificultad se extiende a la *codificación* de las respuestas, cuando aparecen palabras que tienen más de un significado.

Una vez concluido el *trabajo de campo*, a menos que se utilice el *pretest* del cuestionario para el cierre de las preguntas *abiertas*, se procede a la *codi-*

ficación de las respuestas abiertas. Lo habitual es extraer una muestra aleatoria de los cuestionarios completados. La muestra suele oscilar entre el 20 y el 50% del total de cuestionarios (Bourque y Clark, 1994). Varias personas, a ser posible, se encargan de transcribir, literalmente, las distintas respuestas emitidas. Buscarán “términos comunes”, que permitan su agrupación en un número reducido de categorías de respuesta. El *número de categorías* dependerá de la variabilidad de las respuestas y los objetivos de la investigación: si se busca un mayor detalle de respuesta o se prefiere su síntesis en un número reducido de categorías genéricas.

Cada categoría ha de incluir respuestas similares. Su contenido dictará el nombre (o etiqueta) que se dará a la categoría de respuesta. Categorías que han de cumplir los requisitos comunes a la *codificación* de preguntas *cerradas*: exhaustividad, exclusividad y precisión (como después se verá). A cada categoría se le asignará, como en cualquier proceso de medición, un código numérico, que facilitará su análisis estadístico posterior.

- b) *Más expuesta a errores de registro y de traducción de la información*, a una mayor subjetividad en su interpretación. Los entrevistadores (cuando el cuestionario no sea autocumplimentado) han de anotar, literalmente, la respuesta del encuestado, sin introducir ninguna modificación. Ello alarga la duración de la entrevista (comparado con marcar una o varias alternativas de respuesta en una pregunta *cerrada*). Los codificadores también han de proporcionar *etiquetas* que se correspondan con el significado de las respuestas que agrupan. Si las preguntas *abiertas* abarcan un amplio abanico de cuestiones, el análisis de *contenido* es de gran utilidad para el cierre de la pregunta.
- c) *Más susceptibles a errores de medición* debidos a la actuación del entrevistador. Mayor dificultad para alcanzar uniformidad en su actuación: cómo formular las preguntas y anotar las respuestas.
- d) *La respuesta exige más tiempo y esfuerzo por parte del encuestado*. Tanto que su uso se desaconseja en cuestionarios *autoadministrados* (Dillman, 1978). Tener que escribir la respuesta, sobre todo si es detallada y extensa, puede desmotivarle y llevarle a no contestar la pregunta (incrementando la *no respuesta de ítem*) o hacerlo de manera escueta.
- e) *Pueden obtenerse respuestas irrelevantes* o no acordes con la intencionalidad de la pregunta.
- f) *Mayor dificultad para seguir tendencias*. Davis y Smith (1992) destacan la mayor dificultad de mantener reglas de *codificación* útiles a lo largo de los años.

Pese a estas limitaciones, las preguntas *abiertas* son de gran *utilidad*:

- a) En estudios *exploratorios*, cuando no se dispone de un conocimiento previo suficiente del tema que se investiga. A veces puede utilizarse el *pretest* como una primera aproximación exploratoria al tema de estudio y después proceder, a partir de la información recabada, al cierre de la pregunta.

“Quizás el uso más importante de las preguntas abiertas sea pretestar preguntas cuando el investigador quiere explorar muchas dimensiones de un tema y no está seguro de qué preguntas realizar. El uso extensivo de preguntas abiertas con muestras pequeñas puede permitir al investigador desarrollar mejores preguntas cerradas que, cuando sean utilizadas en muestras más grandes, obtendrá los mejores resultados” (Bradburn, 1983: 302).

- b) *Cuando el investigador no prevea todas las posibles respuestas a una cuestión determinada o precise enumerar un listado extenso de respuestas. De especial relevancia en encuestas telefónicas o cuando se quiera ahorrar espacio en el cuestionario.*
- c) *Cuando se busque una respuesta inmediata, prioridades o preferencias.*

Por ejemplo, “En la actualidad, ¿qué es lo que más le preocupa?” o “En la última semana, ¿cuál ha sido la noticia que más ha llamado su atención?”

- d) *Cuando se desee una mayor especificación de una respuesta dada con anterioridad. “Probablemente éste es el tipo de pregunta abierta más habitual en la encuesta” (Payne, 1980: 37). Suele seguir a una pregunta cerrada y se redacta con pocas palabras: “por qué”, “cuáles”, “cómo”. Si bien, se desaconseja su uso reiterado a lo largo del cuestionario para no cansar al encuestado o resultarle impertinente. Conviene no abusar.*

EJEMPLO DE PREGUNTAS ABIERTAS COMPLEMENTANDO A UNA PREGUNTA CERRADA

P11. *¿Recomendaría este centro a un familiar o amigo?*

—	Sí	1
—	No	2
➔	P11a. <i>¿Por qué?</i> _____	

P12. *¿Ha participado Ud. en la organización de alguna de las actividades del centro?*

—	Sí	1
—	No	2
➔	P12a. <i>¿En cuáles?</i> _____	

- e) Si quiere conocerse el valor numérico exacto de una variable.

Por ejemplo, “¿Cuántos años cumplió en su último cumpleaños?” o “¿Cuántas horas o minutos dedica al día a tareas del hogar (excluido el cuidado de los hijos)?”. Grabar el valor numérico exacto amplía las posibilidades de análisis estadístico (frente a su agrupación en intervalos de edad).

En las *encuestas autoadministradas* estas preguntas numéricas pueden aparecer seguidas de una línea o de cuadrados. Los cuadrados tienen la ventaja de indicar el número de dígitos que se solicita.

Mes	Año	
<div style="display: inline-block; border: 1px solid black; width: 30px; height: 20px; margin-right: 5px;"></div> <div style="display: inline-block; border: 1px solid black; width: 30px; height: 20px;"></div>	19 <div style="display: inline-block; border: 1px solid black; width: 30px; height: 20px; margin-left: 5px;"></div> <div style="display: inline-block; border: 1px solid black; width: 30px; height: 20px; margin-left: 5px;"></div>	
¿Cuántos metros cuadrados tiene su vivienda? <div style="display: inline-block; border: 1px solid black; width: 30px; height: 20px; margin-left: 10px;"></div> <div style="display: inline-block; border: 1px solid black; width: 30px; height: 20px; margin-left: 5px;"></div> <div style="display: inline-block; border: 1px solid black; width: 30px; height: 20px; margin-left: 5px;"></div>		

- f) *El tratamiento de temas delicados*, cuando se está interesado en la frecuencia de la conducta. Los encuestados suelen utilizar el rango de alternativas de respuesta numéricas como marco de referencia para estimar la frecuencia de su propia conducta, generando error *sistemático* (Sudman, Bradburn y Schwarz, 1996; Biemer y Lyberg, 2003). Pero, en preguntas relativas al nivel de renta se obtiene una respuesta más sincera en preguntas *cerradas* con amplias categorías de respuesta (Dillman, 1978).
- g) *Para lograr empatía o “rapport” con el encuestado*, ofreciéndole libertad de respuesta.

Por ejemplo, “En su opinión, ¿qué es lo más importante en la educación de los hijos?”.

Preguntas cerradas: también denominadas *precodificadas* o de respuesta fija (porque están ya acotadas en el cuestionario). El encuestado se limita a elegir una o varias (si la pregunta es *múltiple*) de las opciones de respuesta que acompañan a la pregunta. Para evitar errores de medición, conviene que las respuestas se listen verticalmente. A cada respuesta acompaña un número (el *código numérico* que facilitará su transferencia al programa estadístico para su análisis). Además han de cumplir los requisitos de exhaustividad, exclusividad y precisión, exigidos en cualquier proceso de medición (capítulo 2):

- **Exhaustividad:** las categorías de respuesta han de comprender todas las opciones posibles para que ningún sujeto quede sin poder clasificarse. De ahí la recomendación de incluir la opción “otros”, cuando se prevea la posibilidad de otras opciones de respuesta diferentes a las diferenciadas en la pregunta. Y siempre dotándole de espacio suficiente para su especificación.

- **Exclusividad:** las categorías de respuesta han de ser mutuamente excluyentes para evitar solapamientos de respuestas.
- **Precisión:** realizar el mayor número de distinciones posible para obtener una respuesta más precisa. Tiempo habrá para agrupar las diferentes categorías o valores de las variables (en la fase de análisis). En general se observa (Krosnick y Fabrigar, 1997; Fowler y Consenza, 2008) que aumentando el número de categorías hasta al menos siete mejora la calidad de la medición; pero no conviene ir mucho más allá.

EJEMPLO DE PREGUNTAS CERRADAS

Para ilustrar la inclusión de la opción de respuesta abierta “otros” y demás formatos de pregunta cerrada, se toman preguntas incluidas en el estudio del CIS n.º 2.846, la Encuesta CIS-OBEXAXE de *Actitudes hacia la inmigración* de septiembre de 2010, analizada en Cea D’Ancona y Valles (2011). Aparecen como figuran expresas en el cuestionario, exceptuando el tipo de letra (para diferenciar más la pregunta de las opciones de respuesta).

P.14 *¿Y cómo diría Ud. que los españoles, en general, tratan a los inmigrantes extranjeros?* (UNA RESPUESTA) (MOSTRAR TARJETA G).

– Con indiferencia	1	
– Con desprecio	2	
– Con amabilidad	3	
– Con desconfianza	4	(87)
– Con normalidad (igual que si fueran españoles)	5	
– Con agresividad	6	
– N.S.	8	
– N.C.	9	

P.15 *¿Qué influye más en el trato que se da a los inmigrantes en España?* (MOSTRAR TARJETA H).

– Su nacionalidad	1	
– Su cultura	2	
– Su posición económica	3	
– El color de la piel	4	(88)
– Otra respuesta, ¿cuál?		
.....	5	
– N.S.	8	
– N.C.	9	

P.16 ¿Cree Ud. que los inmigrantes instalados en España de manera estable y regular deberían tener el derecho de...?

	Sí	No	NS	NC
– Traer a su familia	1	2	8	9 (89)
– Cobrar el subsidio si se quedan parados	1	2	8	9 (90)
– Votar en las elecciones municipales	1	2	8	9 (91)
– Votar en las elecciones generales	1	2	8	9 (92)
– Crear asociaciones para defender sus derechos	1	2	8	9 (93)
– Obtener la nacionalidad española	1	2	8	9 (94)

P.17 Pensando ahora en la atención sanitaria, ¿está Ud. muy de acuerdo, más bien de acuerdo, más bien en desacuerdo o muy en desacuerdo con las opiniones siguientes?

1. Muy de acuerdo
2. Más bien de acuerdo
3. Más bien en desacuerdo
4. Muy en desacuerdo

	1	2	3	4	NS	NC
– La presencia de inmigrantes hace que disminuya la calidad de la atención sanitaria	1	2	3	4	8	9 (95)
– Los inmigrantes abusan de la atención sanitaria gratuita	1	2	3	4	8	9 (96)
– Los españoles deberían tener preferencia a la hora de acceder a la atención sanitaria	1	2	3	4	8	9 (97)

P.39 Cambiando de tema ¿diría Ud. que, por lo general, se puede confiar en la mayoría de la gente, o que nunca se es lo bastante prudente en el trato con los demás? Por favor, sitúese en una escala de 0 a 10, en la que el 0 significa “nunca se es lo bastante prudente” y el 10 que “se puede confiar en la mayoría de la gente” (MOSTRAR TARJETA M).

Nunca se es lo bastante prudente	Se puede confiar en la mayoría de la gente
00	10
01	
02	
03	
04	
05	
06	
07	
08	
09	

N.S..... 98

N.C..... 99 (161) (162)

P.40 Cuando se habla de política se utilizan normalmente las expresiones izquierda y derecha. En esta tarjeta hay una serie de casillas que van de izquierda a derecha. *¿En qué casilla se colocaría Ud.?* (MOSTRAR TARJETA ESCALA). (PEDIR AL ENTREVISTADO QUE INDIQUE LA CASILLA EN LA QUE SE COLOCARÍA Y REDONDEAR EL NÚMERO CORRESPONDIENTE).
(ENTREVISTADOR: Fíjese que se trata de una escala de 0 a 10 y no de 1 a 10).

Izqda.									Dcha.	
00	01	02	03	04	05	06	07	08	09	10
			N.S.....			98				
			N.C.....			99		(163) (164)		

La última pregunta cumple mejor el requisito de “precisión” que el formato tradicional de variable ordinal con siete categorías de respuesta (extrema izquierda, izquierda, centro izquierda, centro, centro derecha, derecha y extrema derecha) o apenas cinco. Después, en la fase de análisis, se pueden agrupar las categorías inclusive en sólo tres (izquierda, centro, derecha) para la realización de tipologías.

Las preguntas anteriores y sus categorías de respuesta aparecen ordenadas en sentido vertical. Cada opción de respuesta va acompañada de un número, el código que la representa a efectos analíticos. Respecto a la *codificación* se recomienda:

a) Asignar los mismos códigos numéricos a las mismas respuestas.

Por ejemplo, codificar la respuesta “Sí” con el número “1”, “No” con el “2”. Es habitual que las respuestas “no sabe” se codifiquen con los números “8”, “98” o “998”, dependiendo de cuántos dígitos correspondan a la pregunta (de acuerdo con el número de respuestas posibles). La respuesta “no contesta” como “9”, “99” o “999”. En algunas circunstancias puede optarse por reservar el código numérico “0” para designar la *no respuesta* de ítem.

b) El código ha de tener consonancia con el significado de la respuesta. De manera especial en variables *ordinales* (clase social, nivel de estudios, grado de satisfacción), para posibilitar su tratamiento como variable continua.

Por ejemplo, en la variable “clase social”, medida con cinco categorías (alta, media-alta, media, media-baja y baja), los códigos han de ser acordes con el nivel de estatus que designan. El código numérico más bajo, el “1”, para la categoría “clase social baja”; el más elevado, el “5”, para la “clase social alta”; y no a la inversa.

- c) Los *códigos numéricos* suelen figurar *a la derecha de cada opción de respuesta* para facilitar el trabajo del codificador. En los cuestionarios autocumplimentados suele ponerse a la izquierda de la respuesta, para evitar que el encuestado se equivoque. Cuando se pone a la derecha suele quedar espacio en blanco entre la respuesta y su código numérico, pudiendo provocar equívocos, a menos que se pongan puntos suspensivos. En cambio, si el número figura a la izquierda, se deja el espacio a la derecha de las respuestas “libre” por si el encuestado quiere realizar alguna aclaración sobre las respuestas.

En los ejemplos ilustrados también aparecen números entre paréntesis. Corresponden a la *columna* o columnas que ocupará la pregunta en la *matriz de datos*. Cuando la pregunta incluye un número de respuestas inferior a diez, ocupará una sola columna. Excepto que se trate de una *pregunta múltiple*. En ese caso, el número de columnas dependerá del número de opciones de respuesta que se permitan: dos, tres e inclusive todas (si se pide que se ordenen las respuestas de más a menos o de menos a más).

EJEMPLO DE PREGUNTAS CERRADAS MÚLTIPLES

P.11. De las siguientes maneras de invertir dinero, ordénelas de acuerdo con sus preferencias. (Respuesta múltiple). (MOSTRAR TARJETA A). (18-24)

- En cuenta corriente
- En libreta a plazo fijo
- En Deuda Pública o en Bonos Del Estado
- Invertir en Bolsa
- En fondos de pensiones
- Compra de vivienda
- Compra de joyas u obras de arte
- Otras (*Especificar*)

P.12. ¿Qué cualidades, de las siguientes, admira Ud. más en una persona? (Respuesta múltiple). (SEÑALAR SÓLO TRES). (MOSTRAR TARJETA B).

- Lealtad 01
- Respeto 02
- Sinceridad 03
- Tolerancia 04
- Amabilidad 05
- Simpatía 06 (25-26)
- Honestidad 07 (27-28)
- Perseverancia 08 (29-30)
- Humildad 09
- Comprensión 10
- Sencillez 11

Cuando se busque una ordenación, hay que especificarlo en la pregunta indicando la forma de proceder. *Por ejemplo*, escriba “1” junto a la categoría de respuesta que considere más importante; “2”, en la segunda más importante; y así hasta la última. Adviértase que en la primera pregunta, en vez de por códigos numéricos, las respuestas aparecen acompañadas de una línea. Esto es lo recomendable para evitar confusiones entre códigos numéricos.

Cuando la pregunta incluye más de 10 alternativas de respuesta, pero menos de 100, ocupa dos columnas. Si supera los tres dígitos, pero es menor de 1.000, serán tres las columnas asignadas. Y, así, sucesivamente. Asimismo, adviértase que cuando se solicitan dos o más respuestas se está ante una pregunta *múltiple*. En ese caso, hay que dar instrucciones (entre paréntesis, mayúsculas, cursiva o letra diferente a la utilizada para enunciar la pregunta) de cuántas opciones de respuesta se admiten y la manera de indicarlo. Más en el caso de cuestionarios autocumplimentados.

Por ejemplo, “ponga una cruz en la casilla correcta”, “rodee con un círculo la respuesta elegida”, “elija sólo una respuesta”, “señale como máximo tres respuestas”.

Por otra parte, adviértase que cuando la pregunta incluye muchas alternativas de respuesta, el uso de *tarjetas* es habitual en la encuesta *cara a cara*. Las tarjetas permiten visualizar las diferentes respuestas antes de responder la pregunta, mientras que el entrevistador las lee (para evitar molestar al encuestado: problemas de lectura, de visión).

En cuanto a las *opciones de respuesta*, como se dijo en el apartado 6.1, en la respuesta incide qué alternativas se ofrezcan, si se presentan de forma equilibrada, y en qué orden. Los llamados efectos de *primacía*, *recencia* y *aquiescencia* (o tendencia al acuerdo indistintamente de lo que se pregunte):

- El efecto de *primacía* (elección de la primera respuesta) se da más cuando el encuestado lee las opciones de respuesta: encuesta *autocumplimentada*, *cara a cara* con hoja autorellenada, *cara a cara* con tarjetas.
- El efecto de *recencia* (elección de la última) es más probable cuando el entrevistador lee las opciones de respuesta (encuesta *telefónica*, *cara a cara* sin tarjetas).

Los tres efectos acontecen más: conforme se incrementan las opciones de respuesta; en personas de bajo nivel educativo; y cuanto menos interés le suscite el tema sobre el que se pregunta. La *saliencia* es la variable más determinante.

Recomendaciones para evitar errores de medición debidos a las opciones de respuesta:

- a) *Redacción “neutral” de la pregunta*, equilibrada en afirmaciones positivas y negativas (de manera especial para evitar el sesgo de *aquiescencia*).
- b) *Adecuar el número de categorías de respuesta*. Aunque ha de cumplirse el requisito de *precisión*, en la encuesta *telefónica* (al no poderse recurrir a *tarjetas* de respuesta) es preferible no más de tres opciones de respuesta y, como máximo, cinco. Ello no es extensible a los otros modos de administración del cuestionario: más de tres opciones de respuesta es lo ideal. Además se ha constatado que las preguntas dicotómicas son más propiciatorias del sesgo de *deseabilidad social*. Razón por la cual, en la encuesta CIS-OBERAXE de

2008 (Cea D'Ancona y Valles, 2009), y posteriores, se modificó la dicotomía “de acuerdo/en desacuerdo”, aplicada en la encuesta de 2007, a cuatro opciones de respuesta: “muy de acuerdo, más bien de acuerdo, más bien en desacuerdo, muy en desacuerdo” (evitando la opción intermedia “ni de acuerdo ni en desacuerdo”).

- c) *Rotar el orden de lectura de las alternativas de respuesta*, si se prevé que su disposición pueda afectar a la respuesta. Más fácil de aplicar en sistemas informatizados como CAPI o CATI.
- d) *No mencionar opciones de respuesta intermedias*. Son un escape fácil ante una elección difícil, que incomode, o simplemente para no meditar la respuesta.

En experimentos realizados por Kalton, Roberts y Holt (1980), la respuesta intermedia acaparaba entre el 15 y el 49% de las respuestas. Para Converse y Presser (1994), la eligen en torno al 20% de los encuestados, pese a no ser la alternativa de respuesta que escogerían, si no se ofrece en el enunciado de la pregunta. En encuestas del CIS sólo se anotan si la persona la menciona de forma espontánea.

- e) *No explicitar opciones de no respuesta* (“no sabe/no contesta”), salvo que su inclusión se estime de interés para los objetivos de investigación.

Por ejemplo, si quiere recogerse la inexistencia de conocimiento u opinión sobre un tema determinado. La experiencia muestra que éstas constituyen opciones de respuesta recurridas cuando no se quiere pensar o manifestar una respuesta concreta. Este problema se evidencia más en cuestionarios *autoadministrados* (al poderse visualizar como opción de respuesta).

Que las preguntas *cerradas* predominen en los cuestionarios se debe a la conjunción de una serie de *ventajas* que se resumen en el cuadro 6.2:

- a) *Rapidez y comodidad de registro*: más sencillo y rápido que escribir la respuesta de una pregunta *abierta*. Se traduce en ahorro de costes del *trabajo de campo* (tiempo de entrevista) y *edición* de los cuestionarios.
- b) *Mayor estandarización de las respuestas*, ampliando sus posibilidades de comparación.
- c) *Posibilidad de centrar las respuestas* en opciones consideradas relevantes y relacionadas con la cuestión que se pregunta.
- d) *Requieren menos esfuerzo* por parte del encuestado. Más adecuada en personas con problemas de comunicación verbal.

Entre los *inconvenientes* destacan:

- a) *Coartan las opciones de respuesta* y simplifican la información obtenida. Si quieren evitarse preguntas no claramente perfiladas, pueden formularse pri-

mero como preguntas *abiertas* y utilizar el *pretest* del cuestionario para su cierre.

- b) *La respuesta puede estar influida por las alternativas ofrecidas* (y ser acorde a nociones preconcebidas del investigador sobre el tema).

“El encuestador que formula preguntas cerradas puede estar especificando categorías de pensamiento que no son representativas de lo que el encuestado piensa. Con el resultado de que los encuestados confirman nuestro propio marco de referencia sin incluso darnos cuenta” (Dillon, 1990: 141).

Experimentos realizados por Dillon (1990) registran diferentes menciones de respuesta con formatos *abiertos* y *cerrados* de preguntas. *Por ejemplo*, a la pregunta “¿Cuál es el problema más importante al que se enfrenta este país en el presente?”, los encuestados que dijeron “delincuencia” en la pregunta *cerrada* duplicaron los recogidos en la pregunta *abierta*. Con anterioridad, Schuman y Presser (1979, 1996) constataron que a la pregunta “¿Qué preferiría usted más de su trabajo?”, el 60% de los encuestados dieron una respuesta (en la pregunta *abierta*) no incluida en las cinco alternativas de respuesta de la pregunta *cerrada*. “Oportunidad de promoción”, *por ejemplo*, sólo fue mencionada por el 1,8% de los encuestados (cuando la pregunta era *abierta*), mientras que con el formato *cerrado* reunió al 17,2% de los sondeados.

- c) *El diseño de una pregunta cerrada es bastante más laborioso* que el de una pregunta *abierta*. Exige del investigador el previo conocimiento de la realidad que investiga, su delimitación y medición expresa. Ha de decidir cómo formular la pregunta, con qué categorías de respuesta, en qué nivel de medición (nominal, ordinal, de intervalo, de razón o proporción), qué códigos asignar.
- d) No permite al encuestado *especificar el significado de su respuesta*.
- e) Cuando se ofrecen muchas opciones de respuesta, cabe la posibilidad de que el encuestado *no medite las distintas alternativas de respuesta*, sino que elija la primera (*primacía*) o la última opción de respuesta (*recencia*).
- f) *La estandarización de las palabras* (que se consigue igualando la literalidad de la pregunta y sus respuestas) *no implica estandarización de los significados*. A una misma pregunta pueden atribuirse significados diferentes.

Respuestas como, *por ejemplo*, “mucho”, “poco”, “bastante”, “joven”, “caro”, no son igualmente interpretables. Unos pueden considerar que fumar cinco cigarrillos al día es “poco”, mientras que para otros es “mucho”. Una persona de 40 años puede ser “joven” para unos y mayor para otros. Siempre cabe la duda de si las diferentes opciones de respuesta son igualmente interpretadas por todos los encuestados.

- g) *El formato de pregunta cerrado puede incrementar la “amenaza” de la pregunta*, porque “fuerzan” a elegir una de una serie de respuestas. Razón por la

cual Sudman y Bradburn (1974) desaconsejan el formato *cerrado* para preguntas con respuestas socialmente deseables.

CUADRO 6.2. Ventajas e inconvenientes de las preguntas abiertas y cerradas

	<i>PREGUNTAS ABIERTAS</i>	<i>PREGUNTAS CERRADAS</i>
<i>Ventajas</i>	<ul style="list-style-type: none"> – Elaboración sencilla – Ocupan menos espacio – Proporcionan información más amplia y exacta – Pueden obtenerse respuestas inesperadas – Permiten captar lo más “saliente” 	<ul style="list-style-type: none"> – Rapidez y comodidad de registro – Mayor estandarización de la respuesta – Mayor comparabilidad – Respuestas centradas o relevantes – Requieren menos esfuerzo del encuestado
<i>Inconvenientes</i>	<ul style="list-style-type: none"> – Mayor coste de registro, transcripción y codificación – Respuesta más laboriosa – Más expuestas a errores de registro y subjetividad en su interpretación – Pueden darse respuestas no acordes a la pregunta – Mayor dificultad para trazar tendencias 	<ul style="list-style-type: none"> – Coartan las opciones de respuesta – Las alternativas de respuesta marcan la respuesta – Redacción más laboriosa y de mayor conocimiento – No permiten especificar el significado de la respuesta – Posibilidad de respuesta no meditada (<i>primacia, recencia, aquiescencia</i>) – Respuestas de distintas interpretaciones – Pueden incrementar la “amenaza” de la pregunta

6.1.4. La ubicación de la pregunta en el cuestionario

Las preguntas han de estar dispuestas en el cuestionario siguiendo un orden lógico, que reduzca errores de medición. Como ya se dijo en el apartado 6.1, en la calidad de la respuesta no sólo afecta la redacción de la pregunta, sino también el contexto que la comprenda y su ubicación al principio, en el medio o al final del cuestionario. El cuadro 6.3 resume recomendaciones relativas a la *ubicación* de las preguntas para reducir errores de medición. Éstas se detallan a continuación:

CUADRO 6.3. Recomendaciones para la ubicación de las preguntas en el cuestionario

-
- Distribuir secuencialmente las preguntas que formen una *batería*
 - Una pregunta abierta entre dos cerradas
 - Preguntas iniciales interesantes y sencillas
 - Preguntas claves en el centro del cuestionario
 - Preguntas “amenazantes” al final del cuestionario
 - Cuestionario no muy extenso, que no fatigue
-

Las preguntas han de seguir una secuencia “lógica”. El problema está en cómo distribuir las preguntas para evitar el efecto de *contexto*; es decir, la transferencia de significado de una pregunta a la siguiente. Efecto más presente en: preguntas que abordan una misma temática; preguntas *genéricas* (más que en específicas); cuando el encuestado quiere mostrar consistencia en la respuesta a la segunda pregunta; y en personas de bajo nivel educativo. Para su disminución se recomienda:

- a) *Distribuir secuencialmente las preguntas que formen una batería.* Por “batería” se entiende un conjunto de preguntas que abordan una misma cuestión, que se complementan al enfocar diferentes aspectos de un mismo tema. Conviene que las preguntas que formen una *batería* figuren juntas en el cuestionario, formando una unidad. Aunque a veces se prefiera distanciarlas para evitar que sean respondidas buscando coherencia entre ellas. En tal caso se aconseja una breve presentación o introducción a la pregunta, que indique que se pasa a otro tema (*por ejemplo*, “Cambiano de tema...” o “Ahora algunas preguntas sobre...”). Ello puede ayudar a comprender la pregunta y que el cuestionario parezca menos caótico.

Si el tema es complejo y el investigador quiere asegurarse que se consideren diferentes aspectos al responder la pregunta, mejor seguir la técnica del *embudo invertido*: primero formular las preguntas *específicas* y gradualmente ir pasando a *genéricas*. De esta manera se favorece que la respuesta a la pregunta genérica sea más reflexiva, que se valoren diferentes aspectos al responder, además de reducir la *no respuesta*.

Así lo constataron Schuman y Presser (1981/1996): un mayor porcentaje de personas se autodeclaraban “felices”, cuando la pregunta genérica sobre la *felicidad* figuraba en primer lugar (52%), que cuando aparecía después de una pregunta específica sobre la *felicidad conyugal* (38%). Cea D’Ancona y Valles (2009, 2010b) también observan que la *inmigración* obtiene una valoración genérica menos positiva, cuando va después de preguntas específicas sobre inmigración que cuando las precede.

Por el contrario, cuando se sigue la técnica del *embudo* (comenzar con preguntas genéricas, sencillas y menos comprometidas, y gradualmente pasar a preguntas más específicas, complejas y comprometidas), las preguntas *espe-*

cíficas suelen responderse en concordancia con la respuesta dada a la pregunta genérica (Singleton y Straits, 2002).

- b) *Poner una pregunta abierta entre dos cerradas ayuda a reducir el efecto de orden* entre las dos preguntas cerradas (así lo muestran Mason, Carlson y Tourangeau, 1994). Además conviene evitar la fatiga (reducir la monotonía), intercambiando diferentes formatos de preguntas.
- c) *Preguntas iniciales interesantes y sencillas*. Han de cumplir la doble condición de despertar el interés del encuestado y ser, a su vez, sencillas de responder por el conjunto de la población. Ha de evitarse que suscite un elevado porcentaje de *no respuesta*, porque desanimaría a seguir respondiendo el cuestionario.

Mejor preguntas neutras, referidas a aspectos objetivos más que subjetivos, que no pidan decantarse “a favor” o “en contra”, y que estén relacionadas con el problema de investigación. En encuestas mediante *entrevista*, mejor una pregunta *abierta*, genérica, relacionada con el tema central de la encuesta. En cuestionarios *autocumplimentados*, mejor una pregunta *cerrada*. Como indican Sudman y Bradburn (1987: 218), “Las preguntas abiertas, que requieren escribir más de unas cuantas palabras, se perciben como difíciles, como potencialmente embarazosas, debido a la posibilidad de cometer errores gramaticales”. Además, ha de transmitirse la sensación de que contestar el cuestionario va a suponer poco esfuerzo. La decisión sobre qué pregunta poner puede dejarse para el final.

- d) *Preguntas claves en el centro del cuestionario*. Por pregunta “clave” se entiende aquella de especial relevancia para los objetivos de la investigación. Conviene que se formulen después de haber despertado el interés del encuestado, pero, antes de que el cansancio pueda afectar a la respuesta (obteniendo respuestas menos meditadas, y más “no sabe / no contesta”).
- e) *Preguntas “amenazantes” al final del cuestionario*, cuando la reacción a ellas no pueda afectar a la respuesta de preguntas clave. Por ello es práctica común situar las preguntas sociodemográficas (edad, nivel de estudios, ocupación, religiosidad, ideología política, ingresos...) al final del cuestionario. Además es más probable que se aporten datos personales cuando se ha averiguado el propósito de la encuesta. De todas formas conviene que el bloque de preguntas *sociodemográficas* vaya precedido de una *introducción* (o preámbulo), que indique su finalidad en la investigación. Cuantas más características sociodemográficas se recojan, mejor para la explicación del objeto de estudio.

Diseñar un cuestionario no muy extenso, que no fatigue, para que no merme la calidad de las respuestas. La *fatiga* provoca una mayor *no respuesta* de ítem. Por lo que habrá que eliminar toda pregunta repetitiva o no relevante para los objetivos del estudio; y evitar la monotonía de una larga serie de preguntas con un igual formato de respuesta (tipo escala de acuerdo/en desacuerdo).

Por lo general, se recomienda no sobrepasar los tres cuartos de hora (mejor treinta minutos) en encuestas *cara a cara*. En encuestas *telefónicas*, no más de quince minutos; *cara a cara* en la calle, incluso menos (de cinco a diez minutos). En las encuestas *web* incluso hay más presión para que sean breves. No obstante, la fatiga dependerá del: interés que suscite el tema; cómo se haya diseñado el cuestionario; y la destreza del entrevistador.

6.1.5. La forma del cuestionario

Todo cuestionario comienza con un *mensaje de presentación*, que ha de ser cuidadosamente elaborado para incentivar que se responda y se haga de manera veraz. Incluye la identificación del centro responsable del estudio (sobre todo si es de reconocido prestigio); la finalidad de la encuesta, el procedimiento seguido para la selección de encuestados, la información que se precisa, el interés del estudio y los beneficios de la participación. Esto último es muy importante que se resalte. No se olvide que el objetivo es conseguir que la persona participe en la investigación. También afectan las garantías de *confidencialidad* y de *anonimato* dadas. Todo ello de forma escueta. Una presentación larga puede desmotivar la participación (en especial, en la encuesta telefónica). En la encuesta *telefónica* es importante verificar el número de teléfono marcado antes de comenzar la entrevista.

EJEMPLO DE MENSAJE DE PRESENTACIÓN

Mensaje de presentación de la encuesta CIS-OBERAXE de 2010 (estudio n.º 2.846):

“Buenos días/tardes. El Centro de Investigaciones Sociológicas está realizando un estudio sobre temas de interés general. Por este motivo solicitamos su colaboración y se la agradecemos anticipadamente. Esta vivienda ha sido seleccionada al azar mediante métodos aleatorios. Le garantizamos el absoluto anonimato y secreto de sus respuestas en el más estricto cumplimiento de las Leyes sobre secreto estadístico y protección de datos personales. Una vez grabada la información de forma anónima, los cuestionarios individuales son destruidos inmediatamente”.

Mensaje de presentación de la encuesta experimental *MEXEES I* (Cea D’Ancona y Valles, 2010a):

“Buenos días/tardes. Me llamo (...) Soy un profesor/a (alumno/a) de la Universidad Complutense de Madrid. Estamos participando en una investigación sobre las actitudes ante la inmigración. La finalidad del estudio es mejorar su medición mediante encuesta. Motivo por el cual solicitamos su plena colaboración y se la agradecemos anticipadamente. Es muy importante que **participe** y que responda a las preguntas con la mayor **sinceridad** posible para que puedan lograrse los objetivos del estudio. Usted ha sido seleccionado al azar, aplicando métodos aleatorios. Le garantizamos la **plena confiden-**

cialidad y anonimato de sus respuestas, en el más estricto cumplimiento del secreto estadístico. Con su colaboración podremos contribuir a la mejora de la política inmigratoria (No le llevará mucho tiempo responder)".

Estos *mensajes* aparecen impresos en los cuestionarios. El entrevistador tiene que aprenderlos, memorizarlos, pero nunca leerlo literalmente. Sirven de pauta para la presentación de la encuesta.

En la encuesta por *correo*, la introducción aparece en forma de *carta de presentación*, cuya extensión no debería superar una página. Conviene que incluya la fecha de remisión del cuestionario (normalmente, dos semanas después de la fecha de envío), un número de teléfono para cualquier aclaración (se añade al nombre y dirección de la entidad responsable del estudio), e instrucciones básicas para su correcta cumplimentación.

Por ejemplo, en preguntas *cerradas*: ponga una cruz en el recuadro junto a la respuesta elegida; rodee con un círculo el número que corresponda a su respuesta (es preferible el círculo a tachar el número para evitar equívocos). En preguntas *abiertas*: escriba su respuesta en las líneas impresas en la pregunta.

Es habitual incorporar una ilustración gráfica (dibujo o fotografía) para darle apariencia de profesionalidad. También algún incentivo o *gratificación* como estrategia para aumentar la tasa de respuesta (capítulo 7), aunque ésta sea sólo proporcionarles los resultados del estudio. Cuando se trata de una gratificación "prometida" (participar en un sorteo, por ejemplo), en la *carta* se indica el procedimiento para su obtención. La *carta* concluye agradeciendo la colaboración, el saludo cordial y la firma del responsable de la investigación; esta última a ser posible, en bolígrafo azul para darle un toque más personal (pues parece que las cartas se han firmado una a una).

Dillman (2008) recomienda poner en la portada un título. *Por ejemplo*, "Cómo podemos proteger el medio ambiente", más que un título más neutral tipo: "Opiniones sobre temas medioambientales".

Al final del cuestionario es habitual proporcionar un espacio para que el encuestado realice comentarios adicionales sobre el tema de estudio (*por ejemplo*, "Si hay algo más que Ud. desea añadir, por favor, escríbalo en el espacio en blanco a continuación" – si es autocumplimentado), o el cuestionario y su cumplimentación. Si el cuestionario se aplica mediante entrevista, ha de reservarse un espacio para los *datos de control* (precisos para la *supervisión del trabajo de campo*). Las preguntas relativas al procedimiento seguido para seleccionar los entrevistados suelen aparecer al principio del cuestionario, antes del *mensaje de presentación*: n.º de estudio, n.º de cuestionario, comunidad autónoma, municipio, tamaño de hábitat, distrito, teléfono, fecha, nombre del entrevistador, código y datos fijados para las *cuotas* (generalmente sexo y edad) y las *rutras* (si se sigue un muestreo por *rutras aleatorias*). Las preguntas concernientes a la realización de la entrevista figuran al final del cuestionario.

EJEMPLO DE DATOS DE CONTROL

Extracto de la encuesta CIS-OBERAXE de 2010 (estudio n.º 2.846), e igual en las posteriores:

A RELLENAR POR EL ENTREVISTADOR

INCIDENCIAS ENTREVISTA:

- | | | | | |
|--|-------|-------|-------|-------|
| I.1 Número de orden de entrevista (por muestra) | _____ | (226) | (227) | (228) |
| I.2 Dificultad de acceso al edificio, casa, urbanización, etc. | _____ | (229) | (230) | (231) |
| I.3 Viviendas en las que no hay nadie | _____ | (232) | (233) | (234) |
| I.4 Viviendas en las que se niegan a recibir ninguna explicación | _____ | (235) | (236) | (237) |
| I.5 Negativas de varones a realizar la entrevista | _____ | (238) | (239) | (240) |
| I.6 Negativas de mujeres a realizar la entrevista | _____ | (241) | (242) | (243) |
| I.7 Contactos fallidos por no cumplir cuotas | _____ | (244) | (245) | (246) |
| I.8 Contactos fallidos por no ser una vivienda | _____ | (247) | (248) | (249) |
| I.9 Viviendas de inmigrantes | _____ | (250) | (251) | (252) |

ENTREVISTA REALIZADA:

Entrevista conseguida: _____ (calle o plaza) _____ (n.º) _____ (piso) _____ (pta.)

E.1 Fecha de realización: _____

(Día)	(Mes)	(Año)
(253) (254)	(255) (256)	(257) (258)

E.2 Día de la semana que se realiza la entrevista: Lunes 1
Martes 2
Miércoles 3
Jueves 4 (259)
Viernes 5
Sábado 6
Domingo 7

E.3 Duración de la entrevista: _____ (en minutos) (260)(261)(262)

E.4 Hora de realización:	La mañana (9-12)	1	
	Mediodía (12-4)	2	(263)
	Tarde (4-8)	3	
	Noche (8-10)	4	

A RELLENAR EN CODIFICACIÓN**C.1 CUESTIONARIO CUMPLIMENTADO:**

Correcta..... 1 (264)
 Incorrecta..... 2
 ↓
 C.1a MOTIVO: _____ (265) (266)

C.3 RESULTADO FINAL:

Entrevista válida..... 1 (271)
 Entrevista anulada..... 2

 C.4 CODIFICADOR Nº _____ (272) (273)

C.2 VALORACIÓN DE LA INSPECCIÓN:

Entrevista no inspeccionada 1
 ↓
 Inspección telefónica 2 (267)
 Inspección personal 3
 Inspección telefónica y personal ... 4

C.2a Resultado inspección:

Entrevista correcta 1
 Entrevista incorrecta 2 (268)
 No se ha establecido contacto 3
 ↓
 C2b. MOTIVO: _____ (269) (270)

Además de esta información (precisa para la *supervisión del trabajo de campo*), puede ser de interés añadir un espacio en blanco para anotar cualquier observación referida al cuestionario, al desarrollo de la entrevista, al encuestado e incluso a las características de la vivienda (si la entrevista se realiza en el domicilio del encuestado). Información que puede ayudar a la interpretación de los resultados de la encuesta.

Recomendaciones respecto al formato del cuestionario:

- Formato de libro.* Parece más profesional, facilita la lectura, la vuelta de página y evita la pérdida de hojas.
- En papel de buena calidad* (color blanco o pastel). En cuestionarios dirigidos a distintos grupos de población (amas de casa, estudiantes, jubilados...), imprimirlos en colores diferentes puede facilitar su administración. Ayuda a identificar la estructura del cuestionario.
- Impresión atractiva, de fácil lectura* (en letra clara y de tamaño suficiente que permita ser leído con facilidad). También *a doble cara, en dos columnas*, porque abarata los costes de reproducción de los cuestionarios y de envío postal (en encuestas por *correo*). Además, transmite la sensación de brevedad (cuestionario no muy extenso). En encuestas web se desaconseja (Toepoel, Das, van Soest, 2008) poner más de un ítem por página porque influye negativamente en la evaluación de su diseño y en la respuesta (respuestas correlacionadas de ítems que figuran en una misma página).

- d) *Preguntas espaciadas* (en página con márgenes equilibrados), *ordenadas en sentido vertical y numeradas consecutivamente*, con números arábigos (del 1 a n). Las que figuran bajo un mismo cabecero pueden identificarse añadiendo una letra (A, B, C...). Y el número de la pregunta puede destacarse para facilitar su identificación (en cursiva, subrayado, en otro tipo de letra).
- e) *Preguntas cerradas con las opciones de respuesta en sentido vertical y con su código numérico*. Puede adjuntarse un recuadro, si se prefiere que se marque con una cruz la casilla que corresponda a la respuesta. *Por ejemplo*,

P.11. *¿Tiene Ud. hijos cursando estudios universitarios?*

- | | | | |
|-------|--------------------------|----------------------|--------------------------------|
| 1. Sí | <input type="checkbox"/> | o, alternativamente: | 1. <input type="checkbox"/> Sí |
| 2. No | <input type="checkbox"/> | | 2. <input type="checkbox"/> No |

- f) *En preguntas abiertas, proporcionar espacio suficiente para anotar la respuesta*. El espacio suele interpretarse como indicador de la cantidad de información que se desea obtener. Normalmente aparecen líneas (____), aunque puede dejarse un espacio abierto. En preguntas *abiertas numéricas* pueden adjuntarse cuadrados para indicar el número de dígitos que se piden.

Mes
Año

Por ejemplo,

20

- g) *Evitar que las preguntas* (y sus opciones de respuesta) *queden partidas entre páginas*.
- h) *Una pregunta larga no ha de ser seguida por una pregunta breve al final de la página*, para evitar su omisión por error (que se pase de largo).
- i) En encuestas mediante entrevista, *proporcionar instrucciones al entrevistador*, en lugares apropiados a lo largo del cuestionario. Las instrucciones han de identificarse con facilidad, mediante una impresión diferenciada del resto (mayúsculas, cursiva, negrita...).
- j) *En preguntas filtro, dar instrucciones* (literales, mediante flechas, recuadros) *que apunten a la siguiente pregunta*. Es esencial que la instrucción figure inmediatamente después de la respuesta. De este modo es menos probable que el entrevistador (o el encuestado) lo pasen por alto. El *recuadro* es muy deseable porque facilita la visualización de la preguntas a realizar a parte de los encuestados. A los que no corresponda, pasan a las preguntas detrás del recuadro. Es preferible que las bifurcaciones debidas a las preguntas *filtro* aparezcan en una misma página.

EJEMPLO DE PREGUNTAS FILTRO

P.15. ¿Está Vd. jubilado?

Sí	1	(54)
No	2	

SÓLO SI ESTÁ JUBILADO/A

P.16 ¿A qué edad se jubiló Vd.?

_____ (55-56)

SÓLO SI SE JUBILÓ ANTES DE
LOS 65 AÑOS

P.16.A. ¿Cuál fue el motivo de su
jubilación? _____

_____ (57)

P.17 Desde que Ud. se jubiló, ¿se ha visto
afectado por alguna de las siguientes cir-
cunstancias? (LEER CADA UNA).

- | | | | |
|--|---|---|------|
| – Ha perdido al cónyuge | 1 | 2 | (58) |
| – Ha transmitido la herencia
a sus hijos | 1 | 2 | (59) |
| – Ha dejado de consumir
algunos productos | 1 | 2 | (60) |
| – Ha tenido que solicitar
algún tipo de ayuda | 1 | 2 | (61) |
| – Ha vendido su casa | 1 | 2 | (62) |
| – Han disminuido sus
amistades | 1 | 2 | (63) |

P.18. Respecto a la relación con sus vecinos del
edificio, ¿diría Ud. que se lleva bien con...?

Todos	1	
La mayoría de ellos	2	(64)
Unos pocos solamente	3	
Ninguno	4	

➔ **P.18.A.** ¿A qué se debe?

.....
..... (65)

P.19. ¿Ha participado en la organización de al-
guna de las actividades del centro?

Sí	1	(66)
No	2	

➔ **P.19.A.** ¿Cuáles?

.....
..... (67-68) (69-70) (71-72)

➔ **P.19.B.** ¿Por qué no?

.....
..... (73)

k) Preparar el cuestionario para el procesamiento de los datos (ahorra tiempo y dinero). Incluye la *precodificación* de las preguntas cerradas y la *precolumnación* del cuestionario entero (si es necesario), asignando a cada pregunta una o más columnas, dependiendo del número de respuestas posibles. Los *códigos* aparecen entre paréntesis como figura en los ejemplos antes ilustrados.

Finaliza el cuestionario con un “gracias” impreso (para decir al término de la entrevista o para que lo lea el propio encuestado en cuestionarios autoadministra-

dos). Puede también incluirse una breve nota solicitando la revisión del cuestionario (por si ha habido algún error o dejado alguna pregunta sin contestar). En encuestas por *correo* conviene, además, insistir en su rápida devolución.

6.1.6. Tratamiento de temas sensibles y referidos al pasado

El cuadro 6.4 resume recomendaciones para la reducción del *sesgo* de la *deseabilidad social*. Este *sesgo* se infiere cuando hay una sobredeclaración de conductas, opiniones o actitudes que son valoradas de manera positiva, mientras las negativas apenas se declaran. El temor a la desaprobación social lleva al pronunciamiento de respuestas “políticamente correctas” o socialmente deseables, aunque no se correspondan con el sentir real.

CUADRO 6.4. Recomendaciones para reducir el sesgo de deseabilidad social

-
- Motivar al encuestado para que responda con sinceridad
 - Ocultar la intencionalidad de la pregunta
 - Pregunta abierta, larga y sencilla
 - Respuestas flexibles o atenuadas
 - Evitar palabras con reacción estereotipada
 - Aleatorizar las opciones de respuesta
 - Dar mayor privacidad a la respuesta
 - Cargar deliberadamente la pregunta:
 - Todo el mundo lo hace
 - Asumir la conducta y preguntar por su frecuencia y otros detalles
 - Uso de la autoridad para justificar la conducta
 - Razones de por qué no
 - Escoger marcos de tiempo apropiados
 - Ubicar la pregunta al final del cuestionario
-

- a) *Motivar al encuestado para que responda con sinceridad.* Las garantías de confidencialidad y anonimato reducen la “no respuesta” y el sesgo de la deseabilidad social (Singer, von Thurn y Miller, 1995). En la introducción de la pregunta es bueno enfatizar la importancia de proporcionar una información “veraz”, además de garantizar el pleno anonimato de la respuesta.
- b) *Ocultar la intencionalidad de la pregunta.*

Por ejemplo, optar por indicadores indirectos (como los relativos a política inmigratoria) en la medición del racismo y la xenofobia (Cea D’Ancona, 2002b, 2004). Con anterioridad Hesselbart (1975) ya constató que en indicadores “direc-

tos” las personas de mayor nivel educativo son menos propensas a declarar estereotipos raciales; no así en indicadores “indirectos”.

- c) *Pregunta abierta, larga y sencilla.* Formatos de pregunta que incrementan la declaración de conductas sensibles (Bradburn, Sudman *et al.*, 1979; Schuman y Presser, 1981).
- d) *Respuestas flexibles o atenuadas.* Evitar que una respuesta se perciba “correcta” o “incorrecta”. Es habitual relacionar preguntas *amenazantes* con una *no respuesta* elevada.

Por ejemplo, en las encuestas del CIS el porcentaje de *no respuesta* más alto suele darse en las preguntas de *ideología política* e *ingresos*. En la encuesta de 2010 a la que antes se ha hecho referencia (estudio n.º 2.846) los porcentajes de *no respuesta* fueron respectivamente 23 y 29% (de los 2.800 encuestados). Y ello pese a que la pregunta sobre *ingresos* fue la última del cuestionario y en formato flexible:

P.60 Actualmente, entre todos los miembros del hogar (incluido el entrevistado) y por todos los conceptos, ¿de cuántos ingresos netos disponen por término medio en su hogar al mes? No le pido que me indique la cantidad exacta, sino que me señale en esta tarjeta en qué tramo de la escala están comprendidos los ingresos de su hogar.

(**MOSTRAR TARJETA INGRESOS**. Si el entrevistado contesta en pesetas, pasar la cantidad a euros).

– Menos o igual a 300 €.....	01	
– De 301 a 600 €.....	02	
– De 601 a 900 €.....	03	
– De 901 a 1.200 €.....	04	
– De 1.201 a 1.800 €.....	05	(196) (197)
– De 1.801 a 2.400 €.....	06	
– De 2.401 a 3.000 €.....	07	
– De 3.001 a 4.500 €.....	08	
– De 4.501 a 6.000 €.....	09	
– Más de 6.000 €.....	10	
– N.C.	99	

En la pregunta, adviértase la concreción de la variable *renta* (para evitar errores de especificación): “entre todos los miembros del hogar (incluido el entrevistado) y por todos los conceptos”.

- e) *Evitar palabras con reacción estereotipada.*

A las preguntas *directas* (*por ejemplo*, “¿Se considera usted racista?”) hay que concederles una validez limitada en la medición de racismo. Al igual que indicadores *directos* relativos a distancia social (*por ejemplo*: “A usted le preo-

cuparía mucho, bastante, poco o nada que un hijo o una hija suya se casara con un ciudadano de Marruecos u otro país norteafricano?”).

Además se desaconseja el uso de *palabras hipotéticas*, tipo “¿Qué haría si...?” o “¿Le gustaría...?”. Lo que el encuestado dice que “podría hacer” no siempre expresa su conducta futura “real”.

- f) *Aleatorizar las opciones de respuesta.* Una técnica, conocida como *RRT* (*Randomized Response Technique*), originariamente propuesta por Warner (1965); aunque realmente es la pregunta lo que se aleatoriza. Al encuestado se le ofrecen al menos dos preguntas (una “amenazante”, *por ejemplo*, beber, robar..., y otras no). El encuestado elige cuál de las dos preguntas (o más) responde, mediante un procedimiento aleatorio (tirar una moneda al aire, arrojar un dado...). Como las opciones de respuesta a ambas preguntas son las mismas, el entrevistado puede responder “sí” o “no”, sin que el entrevistador sepa realmente qué pregunta está siendo contestada (transmitiendo al entrevistado una mayor privacidad). El inconveniente principal es la dificultad de identificar la respuesta con la persona que la pronuncia (limitando la explicación de lo analizado).
- g) *Dar mayor privacidad a la respuesta.* La autocumplimentación del cuestionario, en su totalidad o en parte (*cara a cara* con hoja de respuesta autorellenada) propicia la captación de respuestas más “sinceras”; menos la encuesta telefónica (Capítulo 5).
- h) *“Cargar” deliberadamente la pregunta.* Se parte de que algunas personas pueden decidir no desvelar conductas, actitudes u opiniones porque creen que se desvían de lo socialmente deseable.

Esta estrategia fue primeramente aplicada por Kinsey y colaboradores (1948), en su famoso estudio de la conducta sexual masculina. Pocos entrevistados admitieron, en una pregunta *abierta*, la práctica de la masturbación (que ellos preveían generalizada). Por lo que optaron por “cargar” la pregunta asumiendo la conducta y preguntando por su frecuencia y otros detalles. *Por ejemplo*, “¿Cuántos años tenía Ud. la primera vez que se masturbó?” (indirectamente se informa que la masturbación es práctica “común”). Para sorpresa de los investigadores, la mayoría de los entrevistados respondió afirmativamente la pregunta.

Junto a ésta, hay otras formas de “cargar” la pregunta. Sudman y Bradburn (1987) las resumen en cinco principales:

- *Todo el mundo lo hace.* Introducir la pregunta con una frase que exprese que se trata de una conducta común o frecuente para reducir la amenaza de su revelación.

Calvo Buezas (2000) hace uso de esta estrategia en su medición de *racismo*, en una encuesta a 6.000 alumnos de primaria y bachillerato, cuando pregunta:

P.11 “**TODOS TENEMOS NUESTROS GUSTOS PERSONALES Y NUESTROS RECELOS. CON ALGUNAS PERSONAS NOS MOLESTARÍA CASARNOS.** *¿Te molestaría casarte con una persona de los siguientes grupos?*”. (Sí o No): moros/árabes, gitanos, judíos, portugueses, negros de África, latinoamericanos, españoles, norteamericanos, europeos, asiáticos, negros de Latinoamérica, indios de América, mestizos, mulatos, blancos.

Adviértase la importancia que se concede a la frase introductoria escribiéndola (en el cuestionario original) en mayúsculas. Sudman y Bradburn (1982) desaconsejan esta práctica en tópicos de amenaza moderada. No así Lensvelt-Mulders (2008: 470), que da el siguiente ejemplo: en vez de preguntar “*¿Alguna vez ha cogido algún equipamiento de oficina que valga más de 25 dólares?*”, que es una pregunta incriminatoria, rephrasearla del modo siguiente: “En todas las oficinas todo el mundo alguna vez se ha llevado a casa algún equipamiento para uso personal. ¿Quién no se ha llevado nunca un lápiz o papel? *¿Alguna vez se llevó a casa equipamiento de oficina que valga más de 25 dólares?*”. Ésta es una pregunta más fácil de responder de manera honesta porque se presenta con un marco más permisivo de referencia.

- *Asumir la conducta y preguntar por su frecuencia y otros detalles*, como se hizo en el estudio de Kinsey y colaboradores (1948).

Por ejemplo, “¿Cuántos cigarrillos fuma al día?”, puede resultar menos estigmatizante que la pregunta: “¿Es usted fumador?”. Inconveniente: las personas que no realicen las actividades sobre las que se pregunta pueden incomodarse ante el supuesto de que sí las practiquen, afectando a su posterior cooperación.

- *Uso de la autoridad para justificar la conducta*. Los encuestados pueden reaccionar de forma más favorable a una aseveración, si se atribuye a alguien a quien respetan o por quien sienten simpatía.

Por ejemplo, “Varios estudios científicos actuales avalan que beber vino reduce la probabilidad de sufrir un infarto *¿Bebe Ud. vino durante las comidas? ¿Con qué frecuencia?*”.

- *Razones de por qué no*. Dar razones que justifiquen la no comisión de conductas socialmente deseables (votar, leer prensa, usar el casco...) reduce la predisposición a no informarlas.

Por ejemplo, “Muchos ciclistas afirman que llevar el casco les incomoda. Pensando en la última vez que Ud. montó en bicicleta o ciclomotor, *¿se puso el casco?*”.

- *Escoger marcos de tiempo apropiados.* Para conductas socialmente no deseables es preferible comenzar con una pregunta como, *por ejemplo*, “¿Alguna vez ha cruzado un semáforo en rojo?”. Si la conducta es socialmente deseable, mejor omitir la expresión “alguna vez” (al encuestado puede molestarle que se dude que realiza conductas socialmente aprobadas).

Por ejemplo, “Pensando en la última semana, ¿cuántos días leyó la prensa?”.

- i) *Ubicar la pregunta al final del cuestionario.* Como ya se dijo antes (subapartado 6.1.4), el objetivo es evitar que la reacción (e inhibición) a la pregunta pueda afectar a la respuesta de preguntas que son claves para la encuesta. Las preguntas sociodemográficas (edad, nivel de estudios, ocupación, religión, ideología política, ingresos...) suelen inhibir al encuestado, al considerarlas personales. De ahí la práctica común de ubicarlas al final del cuestionario. Sobre todo las que más suelen inhibir, como las preguntas relativas al nivel de *ingresos* e *ideología política* (como ya se señaló en el punto d). Pero esta recomendación se extiende a otras preguntas.

EJEMPLO DE UBICACIÓN DE PREGUNTA QUE INHIBE

En la encuesta experimental correspondiente al proyecto MEXEES I (Cea D’Ancona, 2012; Cea D’Ancona y Valles, 2010a), cuyo *trabajo de campo* se realizó del 23 de abril al 20 de mayo de 2008, se estimó conveniente incluir una pregunta donde la persona se autoubicase en una escala de racismo. La pregunta exactamente decía:

P. 32. “Y en la escala de racismo, ¿dónde se ubicaría Ud., siendo 0 ‘nada racista’ y 10 ‘muy racista’?”

Nada						Muy				
00	01	02	03	04	05	06	07	08	09	10

Quería complementarse las respuestas obtenidas a través de indicadores indirectos de racismo y xenofobia. Pero, a su vez, se quería obtener una respuesta con ciertas garantías de validez (ajustada a la realidad). Por lo que se decidió que fuese la última pregunta del cuestionario para evitar el sesgo de la *deseabilidad social*. Además, debía de estar camuflada, de modo que la persona la respondiese con cierta normalidad. Por lo que, se decidió que fuese la última de las preguntas con formato escalar relativas a: la *clase social*, la *ideología política* y la *práctica religiosa*. El interés era analizar aquellos que abier-

tamente se declaraban “racistas”. La pregunta así ubicada resultó un éxito. De los 683 encuestados, sólo 1 no respondió a la pregunta. Una puntuación de 5 o más se obtuvo en el 26% de los encuestados mediante entrevista *cara a cara* y el 29% en la encuesta *autoadministrada*. Además, cuando a continuación se les preguntó, “¿Ha habido alguna pregunta en la que se ha sentido incómodo?”, sólo 8 encuestados (un 1% de la *muestra*) hizo referencia a esta pregunta; 95 encuestados (un 14% de la *muestra*) señaló otras preguntas del cuestionario. Sobre todo las relativas al nivel *educativo* y a la *ideología política*. Las calificaban de preguntas “muy personales”.

El éxito alcanzado en esta pregunta, y a propuesta de la autora, se incorporó a la encuesta que el CIS realiza para OBERAXE en octubre de 2011. Exactamente se propone al CIS su ubicación entre las preguntas que compartan la misma escala de respuesta: de “0” a “10”. Son las preguntas relativas al grado de *confianza en las personas* (P. 39. Cambiando de tema *¿diría Ud. que, por lo general, se puede confiar en la mayoría de la gente, o que nunca se es lo bastante prudente en el trato con los demás?* Por favor, sitúese en una escala de 0 a 10, en la que el 0 significa “nunca se es lo bastante prudente” y el 10 que “se puede confiar en la mayoría de la gente”) e *ideología política* (en una escala de “0” izquierda a “10” derecha). No queda detrás de esta última pregunta porque a la *ideología política* siguen otras preguntas (tradicionales en las encuestas del CIS) referidas al *voto* en las pasadas elecciones generales. A esta última le siguen las demás preguntas sociodemográficas, siendo la pregunta referida al nivel de *ingresos* la última del cuestionario (como en las anteriores encuestas CIS-OBERAXE).

Otra cuestión que ha recibido tratamiento específico, por su repercusión en errores de *medición*, son las *preguntas referidas al pasado*. En general se aconseja evitar preguntas que obliguen a realizar cálculos mentales o recurrir, con frecuencia, a la *memoria*. Estas preguntas ponen en juego la *fiabilidad* de las respuestas e, incluso, la posibilidad de que el sujeto las conteste. En especial, cuando se le pregunta sobre algo trivial o inusual, ha transcurrido tiempo desde su ocurrencia, o se le solicita información detallada. El cuadro 6.5 resume actuaciones habituales ante preguntas referidas al pasado, que se detallan a continuación.

CUADRO 6.5. Recomendaciones para reducir errores de medición por fallos de memoria

-
- Destacar la importancia de responderlas con precisión
 - Preguntas largas
 - Aproximar en el tiempo preguntas de eventos poco notorios
 - Usar procedimientos de ayuda al recuerdo
 - Realizar varias preguntas y situar la de más difícil recuerdo al final
-

a) Anteceder la pregunta con una *introducción que indique la importancia de responder la pregunta con precisión*.

- b) Como en el *sesgo de la discapacidad social*, se prefieren preguntas *largas* porque pueden utilizarse para proporcionar ayudas al recuerdo; llevan más tiempo de lectura (dando más tiempo para pensar y recordar); y el encuestado puede interpretar la longitud de la pregunta como un indicador de importancia, prestándole una mayor atención.

Schwarz *et al.* (2008: 27) recomiendan además el uso del formato de respuesta abierto: “¿Cuántas veces a la semana Ud.....?” _____ veces a la semana”. Aunque las respuestas no sean adecuadas, al menos evitarán sesgos sistemáticos asociados con escalas de frecuencia (nunca, algunas veces o frecuentemente).

- c) *Proporcionar un marco temporal adecuado a la relevancia del tema.* Aproximar en el tiempo preguntas de eventos poco notorios (compra de alimentos, desplazamientos urbanos, programas de televisión vistos...); espaciar aquellos de mayor trascendencia (compra de vivienda, de un vehículo, viajes al extranjero, hospitalizaciones...).

Para Converse y Presser (1994), las preguntas de “promedio” (por término medio), o las relativas a un día típico, suelen ser más útiles que las de un día concreto. Además aconsejan:

- *Estrechar el período de referencia al pasado más inmediato* (la última semana, ayer).

Por ejemplo, “La semana pasada, ¿fue al cine?”. En caso afirmativo, “¿Qué película vio?”.

- *Tomar como referencia acontecimientos o fechas importantes del calendario.*

Por ejemplo, “Desde su último cumpleaños, ¿ha recibido Ud. algún regalo?”.

- *Preguntar por acontecimientos que han sucedido en los últimos seis meses e ir remontándose más allá.* Pero sólo en acontecimientos de cierta notoriedad para el encuestado.

- d) *Uso de procedimientos de ayuda al recuerdo* (Sudman y Bradburn, 1987). Puede consistir en:

- *Proponer ejemplos en la pregunta.*

Por ejemplo, “¿A qué organizaciones pertenece usted? (por ejemplo, religiosas, sindicales, gubernamentales, de caridad, etc.)”.

- *Mostrarles una tarjeta*, que contenga un listado de respuestas, en encuestas cara a cara.

Por ejemplo, en la pregunta “¿Qué hace usted para divertirse?” mostrar una *tarjeta* que comprende distintas actividades lúdicas (ir al cine, ir de compras, dar un paseo, montar en bicicleta, leer, ver la televisión, escuchar música, hacer yoga, darse un baño o ducha, dejando espacio para que se indique otras, si se practican) y preguntar si se realizan o no. Este proceder es frecuente en encuestas de audiencia, hábitos de lectura y actividades lúdicas en general.

- *Mostrar un calendario* con el periodo de tiempo sobre el que se pregunta subrayado (Fowler, 1998).
- e) *Realizar varias preguntas* (referidas a la misma cuestión) *para aumentar la probabilidad de su recuerdo*. Dex (1995/2002) aconseja que, cuando se quiera fechar determinados eventos, aquellos de más difícil recuerdo se pregunten al final de la secuencia de preguntas relacionadas.

6.2. Evaluación del cuestionario

La detección y consiguiente reducción de errores de *medición* en el cuestionario precisa de su precomprobación, al menos mediante el *pretest*. Como indican Esposito y Rothgeb (1998: 566), “Si quieres entender cómo los sujetos responden preguntas de encuesta y cómo diseñar cuestionarios que produzcan datos de alta calidad, identifica preguntas problemáticas e intenta determinar por qué esos problemas existen”. El cuadro 6.6 reúne diferentes procedimientos de evaluación del cuestionario.

CUADRO 6.6. Procedimientos de evaluación de los cuestionarios

-
- Pretest
 - Observación de la entrevista
 - Valoración de los encuestados
 - Valoración de los entrevistadores
 - Consulta a expertos
 - Entrevistas cognitivas: “piensa en alto” y retrospectiva
-

A) *Pretest*

El diseño del cuestionario no ha de darse por concluido hasta que haya sido probado o *pretestado*. Ha de comprobarse su adecuación como instrumento de me-

dición para la consecución de los objetivos del estudio (antes de ser pasado a la muestra de la encuesta). En concreto, se quiere averiguar si:

- a) *Las preguntas tienen sentido, se comprenden y provocan las respuestas esperadas.* La pregunta ha de ser comprendida con facilidad, sin precisar aclaraciones, y tener un mínimo porcentaje de *no respuesta*. Se pueden probar versiones alternativas de una misma pregunta para comprobar cuál satisface mejor el objetivo propuesto.
- b) *Las preguntas cubren la variedad de lo que se analiza.* No han de ser una pauta del “todo o nada” (Frey, 1989: 159). Preguntas que apenas obtienen variabilidad de respuesta (al haber aglutinado a tres cuartas partes o más de los sondeados en una misma alternativa de respuesta), no son “buenas” preguntas. Habrá que analizar el motivo de dicha coincidencia mayoritaria: si se debe a la redacción y al *formato* de la pregunta o a su temática. Los sesgos de *deseabilidad social*, *acquiescencia*, *primacía* o *recencia* se conexionan a la elección mayoritaria de una única respuesta.

En las preguntas *cerradas*, si se prevén problemas en la categorización de las respuestas, es preferible darle un formato inicial de pregunta *abierta* y, tras el *pretest*, cerrarlas.

- c) *La disposición conjunta de las preguntas sigue una secuencia lógica y no influye en la respuesta.* Si se prevén efectos de *orden*, pueden probarse diferentes ubicaciones de las preguntas y comprobar cuál logra respuestas más “veraces”.
- d) *Las instrucciones se entienden.* Al igual que las *transiciones* de preguntas y las *rutas* a través del cuestionario (marcadas por las preguntas *filtro*). Un porcentaje de *no respuesta* elevado puede deberse a una mala redacción de la pregunta, a las opciones de respuesta, a su no administración por “omisión”, por “rutas” no claramente definidas a lo largo del cuestionario.
- e) *La duración del cuestionario no afecta a la respuesta.* Se precisa contabilizar su duración conjunta, registrarse cuando comienza y concluye su administración. Asimismo, es de interés comprobar la duración media de cada sección del cuestionario. Permite averiguar cuál puede ser la responsable de una duración superior a la prevista y si ésta se debe a una dificultad (para responder a la pregunta) mayor de la esperable. Para ello habrá que reservar espacios para anotar el tiempo empleado en su contestación. Más fácil en los métodos asistidos por ordenador (CAPI, CATI, CASI), como se vio en el capítulo 5. Latencias de las respuestas superiores a las esperadas indican problemas en la comprensión de las preguntas, en su enunciado o las opciones de respuesta, que han de llevar a su revisión.

Cuando el *pretest* se ajusta a las condiciones “reales” en las que se aplicará la encuesta, puede además proporcionar información de relevancia referida a:

- a) *Porcentaje aproximado de no respuesta*: problemas para acceder a encuestados, la negativa de éstos a participar.
- b) *Variabilidad de la población* respecto al tema que se analiza. De gran utilidad para el diseño de la muestra (capítulo 7).
- c) *Preparación de los entrevistadores*: permite detectar posibles dificultades para la realización de las entrevistas.
- d) *Coste aproximado del trabajo de campo* (en tiempo y dinero).

Adviértase que la generabilidad de la información que se extraiga dependerá de los casos observados: su número, composición y procedimiento de selección. El *pretest* suele hacerse en una *muestra* pequeña (circa 50 casos), de iguales características a la población de estudio. Lo deseable es que ésta se escoja siguiendo un procedimiento aleatorio, igual que se aplicará en la *muestra* definitiva. Una *muestra* mayor de 100 casos permitirá, además, realizar análisis estadísticos multivariantes, que permitan calibrar la relevancia de las variables incluidas en el cuestionario para explicar el objeto de estudio.

Tal es la importancia del *pretest*, que autores como Sudman y Bradburn (1982/1987: 283) aconsejan que: “si no se tienen recursos para la prueba piloto del cuestionario, no se haga el estudio”. En su opinión, habría que hacer, como mínimo, cuatro pruebas o revisiones de los cuestionarios, antes de su lanzamiento definitivo. Converse y Presser (1994) lo reducen a dos; Oppenheim (1966) lo incrementa a ocho. También es deseable que el investigador no se mantenga al margen, sino que participe en el *pretest* (como entrevistador o supervisor). Cualquier modificación del cuestionario que resulte del *pretest* también conviene precomprobarla antes del lanzamiento definitivo del cuestionario. Si bien, la práctica de prueba continua del cuestionario es inusual porque incrementa los costes de la investigación (en tiempo y dinero).

B) Observación de la entrevista

Acompañar al entrevistador en la realización de la entrevista (o escucharla, en el caso de la encuesta *telefónica*), e incluso grabarla (con autorización del encuestado), es de gran ayuda para detectar incorrecciones en su realización y posterior mejora. En la *encuesta telefónica* la *grabación* es fácil; en la *cara a cara* también, aunque con la grabación digital conjunta con CAPI (CARI: *Computer-Audio-Recorded Interviewing*). La *observación* puede hacerse en la fase de *pretest* y cuando se administre el cuestionario definitivo. Entre los indicios de “fallos” o problemas en el cuestionario están:

- a) Si el entrevistador *no lee la pregunta exactamente como está escrita* puede deberse a que la pregunta esté mal redactada.
- b) Si el entrevistador *omite alguna pregunta*, puede deberse a especificaciones incorrectas de las preguntas *filtro*, sus *rutas* o ubicación de la pregunta (*por ejemplo*, una pregunta corta que sigue a una larga).

- c) Si el entrevistado *solicita alguna aclaración* sobre la pregunta, es indicativo de que la pregunta no es “clara”: problemas de definición o de comprensión de su utilidad o intencionalidad.
- d) Si el entrevistado *interrumpe al entrevistador*, antes de que éste acabe de leer la pregunta, puede indicar que la pregunta es demasiado larga.
- e) Si el entrevistado *tarda más tiempo* (del esperable) *en responder*, se debe a problemas en la pregunta: comprensión de su enunciado y de las opciones de respuesta.

Bassili y Scott (1996) constatan que preguntas con dobles negativas tardan más tiempo en responderse que cuando se redactan en afirmativa.

- f) Si el entrevistado proporciona *respuestas “inadecuadas”*, también es indicativo de “fallos” en la pregunta.

C) Valoración de los encuestados

La valoración de quien cumplimenta el cuestionario, los encuestados, es de gran utilidad en su evaluación. La valoración puede obtenerse de forma individual o colectiva (en grupos):

- a) Lo más sencillo es reservar un espacio, al final del cuestionario, para incluir preguntas relacionadas con el cuestionario: interés, extensión, grado de dificultad de la respuesta (en general y en preguntas concretas, que se prevean de especial dificultad), aclaraciones de determinadas respuestas y “mejoras” que introducirían en el cuestionario (redacción de las preguntas, su localización y la presentación, en general, del cuestionario).
- b) Realización de grupos de discusión con encuestados, tras la conclusión del *trabajo de campo* (complimentación de los cuestionarios) o del *pretest*. En este último caso pueden arbitrarse actuaciones que mejoren la aplicación definitiva del cuestionario.

D) Valoración de los entrevistadores

De igual relevancia es la valoración de los entrevistadores. Ésta puede asimismo recabarse de manera individual y colectiva:

- a) Lo más sencillo es proporcionar un espacio, al final del cuestionario (en el apartado correspondiente a *datos de control*), para que los entrevistadores anoten cualquier incidencia que hayan observado durante la realización de la entrevista. La finalidad es detectar cualquier “problema” que se haya dado durante la aplicación del cuestionario, en general, y de preguntas concre-

tas, en particular. Además puede solicitárseles propuestas dirigidas a mejorar el cuestionario (para estudios posteriores).

EJEMPLO DE VALORACIÓN DEL ENTREVISTADOR

Datos correspondientes a la realización de la entrevista de la encuesta CIS-OBERAXE de 2010 (estudio n.º 2.846), que ha de cumplimentar el entrevistador:

A RELLENAR POR EL ENTREVISTADOR

P.62 VALORACIÓN DE LA ENTREVISTA:

- Se ha realizado la entrevista en presencia de terceras personas 1 (203)
- Ha expresado el entrevistado deseo de abandonar la entrevista antes de finalizarla 1 (204)
- Se ha sentido el entrevistado incómodo o molesto por el tema de la encuesta 1 (205)
- Ha tenido prisa el entrevistado por acabar la entrevista 1 (206)

P.63 ¿Ha habido alguna pregunta concreta que provocara incomodidad? (ANOTAR N.º DE PREGUNTA. MÁXIMO 5).

(207) (208) (209)	(210) (211) (212)	(213) (214) (215)
(216) (217) (218)	(219) (220) (221)	

P.64 ¿Han intervenido activamente terceras personas en el desarrollo de la entrevista?

- Sí 1 (222)
- No 2

P.65 Respecto a las tarjetas...

- La persona entrevistada las ha usado todas 1
- Sólo ha usado algunas 2 (223)
- Las he leído yo 3

P.66 Desarrollo de la entrevista:

- Muy buena 1
- Buena 2
- Regular 3 (224)
- Mala 4
- Muy mala 5

P.67 Sinceridad del entrevistado:

- Mucha 1
- Bastante 2 (225)
- Poca 3
- Ninguna 4

b) De manera complementaria a la información individual anterior, resulta muy enriquecedor si, inmediatamente después del *pretest*, se reúne a varios de los entrevistadores. Es preferible que haya proximidad en el tiempo para favore-

cer el “recuerdo”. Al recuerdo contribuyen los datos de *control* (antes referidos) y las impresiones que el entrevistador registre durante la entrevista. En la puesta en común los entrevistadores aportarán información sobre las incidencias habidas durante la administración del cuestionario, tanto relativas al cuestionario como al “campo” (contactar y “retener” a los entrevistados).

E) Consulta a expertos

Puede hacerse de manera individual o en grupo. La finalidad es identificar errores de *especificación* y de *medición* debidos al cuestionario. Por su condición de “expertos” (personas con experiencia en la práctica de encuesta o en el tema sobre el que versa el estudio), se les solicita que analicen la calidad del cuestionario:

- a) Que revisen las preguntas, relacionándolas con los conceptos a medir: su redacción, si inducen a respuestas socialmente deseables y demás errores de *respuesta*.
- b) Su ubicación en el cuestionario y presentación. Un énfasis especial se pone en las *instrucciones* y las *rutras* a lo largo del cuestionario.

Para obtener una información más reflexiva, es mejor enviarles el cuestionario una semana antes de la fecha fijada para la entrevista individual o la reunión de grupo. Sus comentarios se graban en cintas de audio e, incluso, de vídeo, siguiendo el proceder habitual en cualquier grupo de discusión y entrevista cualitativa.

Lo más económico es solicitar la información de forma individual y por escrito. Lo más enriquecedor, aunque de mayor coste, es la discusión en grupo.

F) Entrevistas cognitivas

Tradicionalmente aplicadas por los psicólogos cognitivos para conocer el proceso de pregunta y respuesta. Pero en las últimas décadas se han convertido “en un componente importante de las actividades de precomprobación de la encuesta” (Biemer y Lyberg (2003: 264). Suelen realizarse en espacios “controlados”, equipados con grabadoras y vídeo (a modo de laboratorio); aunque también pueden llevarse a cabo en el domicilio del entrevistado, e incluso por teléfono. En cualquiera de las situaciones, la entrevista se graba para ser, con posterioridad, analizada con detenimiento.

La *entrevista cognitiva* puede tener un carácter simultáneo o retrospectivo. La primera es la comúnmente conocida como entrevista “*piensa en alto*”, en tanto que la segunda como entrevista *retrospectiva*:

- a) *Simultánea* o “*piensa en alto*”: se pide a la persona que diga, en voz alta, lo que piensa cuando se le formula la pregunta y medita la respuesta.

- b) *Retrospectiva*: después de haber respondido al cuestionario, se le solicitan “acleraciones” de algunas respuestas.

Por ejemplo, “¿Qué pensó cuando la pregunta _____?”; “¿Por qué fue la respuesta elegida _____ y no otra?”. La ventaja es que no introduce *sesgos* en la respuesta a preguntas posteriores, además de no alargar en exceso la entrevista (a diferencia de la *simultánea*). El inconveniente principal es el “olvido”: como la información no se transmite a la vez que se produce, la persona puede no recordar el proceso cognitivo seguido en la respuesta a las preguntas del cuestionario.

Ambas variedades muestran gran utilidad para conocer el proceso de respuesta y para detectar errores por problemas de comprensión de la pregunta, de retención de la información (que se dé en la pregunta) y las opciones de respuesta. En su contra está que se limite al conocimiento de la reacción que suscitan las preguntas del cuestionario (no aportando otra información de relevancia para el correcto desarrollo de la encuesta). Por ello, se recomienda que su aplicación sea complementaria a otras “pruebas” del cuestionario. Y tras cualquier modificación del cuestionario, su precomprobación antes de la aplicación definitiva del mismo.

Lecturas complementarias

- Campanelli, P. (2008): “Testing survey questions”, en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Taylor & Francis Group. Nueva York: 176-200.
- Dillman, D. A. (2008): “The logic and psychology of constructing questionnaires”, en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Taylor & Francis Group. Nueva York: 161-175.
- Dillman, D. A.; Smyth, J. D. y Christian, L. M. (2009): *Internet, mail and mixed-mode surveys. The Tailored Design Method*. John Wiley & Sons, Inc. Hoboken, Nueva Jersey.
- Fowler, F. J. Jr, y Cosenza, C. (2008): “Writing effective questions”, en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Taylor & Francis Group. Nueva York: 136-160.
- Kreuter, F.; Presser, S. y Tourangeau, R. (2008): “Social Desirability Bias in Cati, IVR, and Web Surveys. The Effects of Mode and Question Sensitivity”. *Public Opinion Quarterly*, 72 (5): 847-865.
- Lensvelt-Mulders, G. (2008): “Surveying sensitive topics”, en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Taylor & Francis Group. Nueva York: 461-478.
- Lozar Manfreda, K. (2008): “Internet surveys”, en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Taylor & Francis Group. Nueva York: 264-284.
- Presser, S.; Couper, M. P.; Lessler, J. T.; Martin, E.; Martin, J.; Rothgeb, J. M. y Singer, E. (2004): “Methods for testing and evaluating survey questions”. *Public Opinion Quarterly*, 68 (1): 109-130.

- Schwarz, N.; Knäuper, B.; Oyserman, D. y Stich, Ch. (2008): “The psychology of asking questions”, en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Taylor & Francis Group. Nueva York: 18-34.
- Schuman, H. y Presser, S. (1996): *Questions & answers in attitude surveys. Experiments on question form, wording and context*. Sage. Thousand Oaks, CA.
- Sudman, S. y Bradburn, N. M. (1982/1987): *Asking questions: a practical guide to questionnaire design*. Jossey-Bass. San Francisco.
- Toepoel, V.; Das, M. y Van Soest, A. (2008): “Effects of design in web surveys. Comparing trained and fresh respondents”. *Public Opinion Quarterly*, 72 (5): 985-1007.

Ejercicios propuestos

1. Indique los diferentes errores debidos al diseño del cuestionario y qué actuaciones son mejores para evitarlos.
2. Diseñe un cuestionario propio sobre *discriminación*. El cuestionario ha de estar codificado y precolumnado. Exponga las precomprobaciones que haría antes de la redacción definitiva del cuestionario y su justificación.
3. Las siguientes preguntas se incluyeron en una encuesta a centros municipales de la tercera edad (realizada en 1993 por M.^a Á. Cea D’Ancona y M. Valles Martínez. Financiada por el Área de Servicios Sociales del Ayuntamiento de Madrid). De ellas, indique cuántas preguntas realmente incluye y cómo se precolumnaría. Observe que es una selección de preguntas. Se extractan tal y como figuran en el cuestionario.

P.18. ¿Cómo conoció este centro? A través de...

Vecinos	1
Familiares	2
Amigos	3
Parroquia	4
El Ayuntamiento (JMD)	5
Otros (especificar)	

P.20. Por favor, ¿recuerda qué días de la semana pasada fue Ud. al centro?

L M X J V S D

P.20.A. (EN CASO DE NO HABER IDO NINGÚN DÍA) ¿A qué se debió? _____

P.20.B. (EN CASO DE HABER IDO ALGÚN DÍA) *¿Cuáles son las tres razones principales por las que acude?* _____

P.24. Los centros municipales de la tercera edad suelen contar con una serie de servicios para sus socios. Dígame si conoce o no en su centro los que le cito a continuación. (EN CASO AFIRMATIVO) *¿Querría calificarlos de 1 a 10, según la opinión que tenga de ellos? (SI VALORA MENOS DE 5) ¿Por qué?*

SERVICIOS	CONOCE		UTILIZA		VALORA	POR QUÉ
	Sí	No	Sí	No		
Podología	1	2	1	2	_____	
Peluquería	1	2	1	2	_____	
Comedor	1	2	1	2	_____	
Cafetería	1	2	1	2	_____	
Gimnasia	1	2	1	2	_____	
Biblioteca	1	2	1	2	_____	
Taller de pintura	1	2	1	2	_____	

P.40. Hablando de su familia, *¿tiene Vd. hijos, hermanos, sobrinos o nietos? ¿Dónde residen habitualmente?*

	EN CASA		EN MADRID		FUERA DE MADRID	
	Sí	No	Sí	No	Sí	No
Hijos	1	2	1	2	1	2
Hermanos	1	2	1	2	1	2
Sobrinos	1	2	1	2	1	2
Nietos	1	2	1	2	1	2

7

LA SELECCIÓN DE UNIDADES DE OBSERVACIÓN: DISEÑO MUESTRAL Y TRABAJO DE CAMPO

La selección de las *unidades* de observación es una fase esencial de cualquier investigación. De cómo sea su materialización dependerá la calidad de la información que se recoja: las posibilidades de generalización de los resultados del estudio; razón por la cual no hay que descuidar recursos ni en la planificación ni en la ejecución del diseño muestral. Se puede haber proyectado una buena *muestra* (que cumpla criterios de *calidad*), si al final no se consigue que en la investigación acaben participando aquellas unidades inicialmente seleccionadas, se puede acabar ofreciendo una fotografía distorsionada de la realidad. A lo largo del presente capítulo se van a ir desgranando los diferentes integrantes del *diseño muestral* y su puesta en práctica durante el *trabajo de campo*.

7.1. Fundamentos de muestreo

Tres pilares claves en la configuración de la teoría del *muestreo*, y en su aplicación a la investigación social, la proporcionan Kiaer, Bowley y Neyman:

- Anders Nicolai Kiaer (director fundador del instituto noruego de estadística: *Statistisk Sentralbyrå*) fue quien primero abogó por el uso de “*muestras representativas*”. En su libro sobre el método representativo de las encuestas estadística, *Den repraesentative undersøgelsesmethode* (publicado en 1897 en la editorial Kristinia, de Oslo) defiende que se puede extraer información de una *muestra* relativamente pequeña de la población, sin necesidad de proceder a un censo completo de la población (como hasta entonces se creía). Pero ello exige que la *muestra* se seleccione siguiendo algún procedimiento aleatorio que garantice su “representatividad” (que constituya una representación a pequeña es-

cala de la población de estudio), y que de la información extraída de ella pueden estimarse los correspondientes parámetros poblacionales. Dos años antes, en 1895, había aplicado, por primera vez, el término “*investigación representativa*”, en un informe que realiza para el Instituto Nacional de Estadística (que centralizaba el debate de los estadísticos de finales del siglo XIX).

A Kiaer también se debe la introducción del procedimiento de “*replicación*” para evaluar los datos de encuesta: repetir la operación de muestreo, extraer una serie de submuestras comparables. Si bien el método de la *replicación* adquirirá un mayor desarrollo de la mano del estadístico hindú Prasanta Chandra Mahalanobis, en dos artículos publicados en 1944 (“On the large scale sample surveys”. *Royal Society Phil. Trans. B.*, 231: 329-451) y 1946 (“Recent experiments in statistical sampling in the Indian Statistical Institute”. *Journal of the Royal Statistical Society*, 19: 325-378). En ellos aboga por el uso de la *replicación*, “interpenetrar muestras”, para medir la precisión de las estimaciones derivadas de una encuesta. En el último artículo además enfatizó la necesidad de incluir los “*errores humanos*” (la variabilidad introducida por entrevistadores, codificadores y supervisores), junto a los de muestreo, en la estimación de la precisión de una encuesta.

- La propuesta de Kiaer de aplicar la *estratificación* para la selección aleatoria de la *muestra* fue con posterioridad concretada por Jerzy Neyman, en un artículo que publica en 1934 (“On the different aspects of the representative method: the method of stratified sampling and the method of purposive selection”. *Journal of the Royal Statistical Society*, 97: 558-606). En él desarrolla la *estratificación óptima*, el *muestreo por conglomerados*, el *muestreo en poblaciones finitas*, la *distribución muestral* y el *error de muestreo* (o variabilidad de las estimaciones muestrales), que puede medirse a partir de la varianza del estimador.
- Arthur L. Bowley (el primer profesor de estadística en la *London School of Economics and Political Science*) fue quien introdujo el “*muestreo probabilístico*” en la práctica de encuesta. Es la selección “aleatoria” de las unidades que compondrán la “muestra” lo que asegura su “representatividad”, su equivalencia con la *población* de la que se ha extraído. Todos los integrantes de la *población* (objeto de estudio) han de tener igual *probabilidad* de ser elegidos para constituir la *muestra*. No es preciso que ésta sea demasiado grande, sólo que sea “representativa”. Ello permitirá que de las *estimaciones muestrales* se infieran los correspondientes *parámetros poblacionales*, aunque dentro de unos *intervalos de confianza* concretos, como posteriormente precisó Neyman en 1934.

A Bowley se le atribuye la primera encuesta con *selección aleatoria* de los informantes. Una investigación que realiza junto con A. R. Burnett-Hurt, financiada por organizaciones de beneficencia, y que publica en 1915 (*Livelihood and Poverty*. Ed. Bell. Londres). Si bien, en publicaciones anteriores ya había defendido el uso de *muestras aleatorias representativas* en la investigación mediante encuesta: en 1906 (“Presidential address to the Economic

Section of the British Association for the Advancement of Science”, *Journal of the Royal Statistical Society*, 69: 540-558), 1913 (“Working class households in Readings”, *Journal of the Royal Statistical Society*, 76: 672-691) y 1915 (*The Nature and Purpose of the Measurement of Social Phenomena*. P. S. King and Son, Ltd. Londres). En este último artículo definió parámetros de calidad de los datos de encuesta.

Las aportaciones de estos estadísticos fueron con posterioridad desarrolladas y ampliadas en notorios manuales de *muestreo*, entre los que destacan los publicados por:

- F. Yates (1949): *Sampling Techniques of Censuses and Surveys*. Griffin. Londres (tuvo gran influencia en la *encuesta mundial de población* de 1950).
- W. E. Deming (1950): *Some Theory of Sampling*. Wiley. Nueva York.
- M. H. Hansen, W. N. Hurwitz y W. G. Madow (1953): *Sample Survey Methods and Theory*. Wiley. Nueva York (diferencia errores variables y sistemáticos, y propone el concepto de *error total de encuesta*).
- W. G. Cochran (1953): *Sampling Techniques*. John Wiley & Sons. Nueva York. Para O’Muircheartaigh (1997: 7) es el manual de mayor éxito, “el más matemático y menos influido por intereses no estadísticos y no de muestreo”.
- P. V. Sukhatme (1953): *Sampling Theory of Surveys with Applications*. The Indian Society of Agricultural Statistics. Nueva Delhi.
- L. Kish (1965): *Survey Sampling*. John Wiley & Sons. Nueva York (de gran repercusión en España).

En todos ellos se mide el efecto del diseño *muestral* en la imprecisión de las estimaciones de encuesta. Aunque analizar una “muestra”, y no la *población* en su totalidad, permite un mayor control del proceso de investigación (aparte del menor coste económico y temporal), en su contra está el *error muestral*. Por muy “perfecta” que sea la *muestra*, como únicamente se analiza una parte de la población (y esa *muestra* tan sólo constituye una de todas las posibles muestras que pueden extraerse de una misma población), siempre habrá alguna divergencia entre las *estimaciones muestrales* y los correspondientes *parámetros poblacionales*. Pero el *error de muestreo*, aunque fundamental, es sólo uno de los integrantes del *error total de encuesta* (como se vio en el capítulo 5). De él dependen las posibilidades de inferencia, de generalización de los resultados del estudio. Del *error muestral* y sus determinantes se habla en el apartado 7.4, después de haber detallado las distintas fases que comprende el diseño *muestral*.

7.2. Decisiones claves en el diseño muestral

Todo diseño muestral comprende una serie de decisiones claves. Atañen a la *población* (o universo de estudio), al tamaño de la *muestra* y al procedimiento seguido

para su selección (los diferentes procedimientos de selección de las unidades de la *muestra* se concretan en el apartado 7.3).

7.2.1. Delimitación de la población de estudio

La primera decisión concierne a la *población* o universo de estudio. Afecta a las posibilidades de inferencia de los datos del estudio. Además actúa como criterio de evaluación: en qué medida la *muestra* logra ser una representación, a pequeña escala, de la variedad de unidades que componen la *población*. No obstante, esta delimitación no siempre se hace expresa en el informe de la investigación. Como señala Sudman (1983: 150), “Desafortunadamente, los investigadores frecuentemente olvidan hacer explícito el universo que desean estudiar o asumen que el universo corresponde a la muestra seleccionada”.

Por *población* (o *universo*) de estudio comúnmente se entiende el “conjunto de unidades para las que se desea obtener cierta información” (Sánchez Crespo, 1971: 11). Las *unidades* pueden ser personas, familias, viviendas, escuelas, organizaciones, artículos de prensa e, inclusive, acontecimientos (nacimientos, fallecimientos, migraciones). La *información* vendrá dictada por los *objetivos* de la investigación (la integración social de los inmigrantes, la convivencia en el hogar, la valoración de las campañas electorales). El tema y objetivos de la investigación determinan la *población* a analizar. La *definición* de la población ha de incluir:

- a) *Las unidades que la componen*: personas, hogares, facultades, periódicos... La *población* no siempre coincide con la fuente de información.

Por ejemplo, en una encuesta a directores de centros penitenciarios para solicitarles información de sus centros, la *población* de estudio no son los directores, sino los centros que dirigen. Una encuesta a hogares para conocer la situación laboral de sus miembros (a partir de la información dada por uno de sus miembros), la *población* son los hogares, no los individuos. Al igual que una encuesta para conocer la calidad de vida de los enfermos de Alzheimer, la *población* son los enfermos de Alzheimer, aunque las personas que respondan a la encuesta sean las personas que les cuidan (familiares, responsables de los centros de día, personal de residencias de la tercera edad).

- b) *Características sociodemográficas que la distingan de otras poblaciones*: sexo, edad, estado civil, etnia o cualquier característica que esté relacionada con los objetivos del estudio.
- c) *Su ubicación en el espacio y el tiempo*. Por ejemplo, en España, en otoño de 2011.

EJEMPLO DE DEFINICIÓN DE LA POBLACIÓN DE ESTUDIO

- A) Definición de la población de estudio en la encuesta de *Actitudes hacia la inmigración (IV)*, que el CIS (Centro de Investigaciones Sociológicas) realiza para OBERAXE (Observatorio español de racismo y xenofobia) en septiembre de 2010 (estudio CIS n.º 2.846), como aparece en la ficha técnica:

Universo:

Población de nacionalidad española de ambos sexos de 18 años y más

Al ser una encuesta de ámbito nacional, estaría bien que se hubiese explicitado si incluye las provincias insulares (Canarias y Baleares), Ceuta y Melilla. *Por ejemplo,*

Población española, de ambos sexos, de 18 años y más, excluyendo a los residentes en Ceuta y Melilla.

- B) Un estudio para conocer las expectativas laborales de los jóvenes universitarios tiene que concretar si su *ámbito* es *europeo* (va a sondearse a estudiantes universitarios de todos los países de la Unión Europea), *nacional* (estudiantes de universidades españolas), *autonómico* (de una comunidad autónoma), o de una sola universidad. Esta especificación es necesaria porque los resultados del estudio sólo podrán inferirse a dicho universo. Si únicamente participan estudiantes de la Universidad Complutense de Madrid, *por ejemplo*, los resultados de la encuesta no podrán ni siquiera inferirse al conjunto de los universitarios madrileños porque no han participado estudiantes de otras universidades de Madrid (públicas y privadas; habría que especificarlo). Únicamente serán inferibles a los estudiantes de la Universidad Complutense y en una fecha concreta (*por ejemplo*, en el curso 2011-2012) y en los límites que marquen los diferentes *errores* de la encuesta (no sólo los *errores muestrales*, como se vio en el capítulo 5).

La delimitación de la *población* estará determinada por el *problema* de investigación, sus *objetivos* concretos, los *recursos* disponibles (económicos, materiales, humanos), el *plazo temporal* para la realización del estudio y, en el caso de la encuesta, el método a aplicar. La definición de la *población* ha de ser operativa, “factible”, y ajustada a los recursos disponibles. Sus límites (las unidades de la población excluida) condicionarán el alcance de los resultados, sus posibilidades de *inferencia*. Si bien, ha de advertirse que la *población* inicial no siempre coincide con la finalmente sondeada. Depende de la magnitud y las características de los *errores de muestreo*, *cobertura* y *no respuesta*, como se verá a lo largo del capítulo.

7.2.2. Elección del marco muestral y su traducción en errores de cobertura

Delimitada la población, el siguiente paso es la elección del *marco muestral*: el “listado” que incluye a la población de estudio y que se utilizará para la extracción de la *muestra*. Pueden ser datos censales (el *censo de población*, el *padrón de habitantes*, el *censo electoral*), un registro de individuos, de empresas, de pacientes, o cualquier procedimiento que lleve a la identificación de los miembros de la *población* de interés. *Por ejemplo*, el listado de alumnos matriculados en universidades españolas.

A veces el *marco* es conceptual más que físico. Caso de la *encuesta telefónica*, cuando la selección de la *muestra* (los números de teléfono) se realiza no mediante las guías telefónicas (*marco muestral*), sino a través de la generación aleatoria de números de teléfono: el método *RDD* (*Random-Digit-Dialing*), que genera, de manera aleatoria, un número determinado de dígitos (7, 8, 9..., dependiendo del área que abarque el estudio). El *marco* incluye todos los números de X dígitos, que constituyen números de teléfono potenciales (*por ejemplo*, 913942671, 952281420). El problema principal es no saber a dónde se llama (una vivienda, una oficina, empresa o comercio). Información que sí se tiene cuando se utiliza el listín telefónico como *marco muestral*.

Del *marco* se espera que sea un descriptor válido de la población de estudio. Razón por la cual ha de cumplir una serie de *requisitos* mínimos:

- a) Ha de ser lo más *completo* posible en aras de la *representatividad* de la *muestra*. Ésta sólo podrá “representar” a la *población* que incluya el *marco* (a quiénes se da oportunidad de participar en el estudio).
- b) La comprensividad del marco lleva a la exigencia de su *actualización*, que incluya a los miembros “reales” de la *población*. Se precisa conocer la periodicidad de sus actualizaciones.

Por ejemplo, la razón principal del fracaso del Instituto de la Opinión Pública estadounidense de Gallup, en la predicción de las elecciones presidenciales de 1948 (a diferencia de las precedentes de 1936, 1940 y 1944), se atribuyó (Babbie, 2000) a la no adecuación del *marco muestral*. Como en anteriores sondeos, se realizó un *muestreo por cuotas* a partir del *Censo de Población*. El consultado fue el más reciente, aunque databa de 1940. En 1948 se produjeron cambios notorios en la composición de la población norteamericana, como consecuencia de la Segunda Guerra Mundial. Hubo un desplazamiento masivo de habitantes del campo a las ciudades. Ciudades en las que se registró un voto preferentemente demócrata. Por lo que, la sobrerrepresentación de votantes rurales en la *muestra* afectó a la subestimación del voto demócrata.

- c) Cada unidad de la población ha de tener *igual presencia* en el *marco muestral* (la presencia marca la probabilidad de elección).

Por ejemplo, en encuestas telefónicas a hogares, aquellos con más de un número de teléfono tienen mayor probabilidad de ser seleccionados.

Las *duplicidades*, junto con las *omisiones* de unidades de la población y las *inclusiones* erróneas (unidades que no deberían estar incluidas), son los motivos principales de *errores de cobertura* debidos al *marco muestral*. La *duplicidad* es frecuente cuando se combinan listados diferentes.

d) *El marco no ha de incluir a unidades que no se ajusten a la población de estudio.*

Por ejemplo, si la población son jóvenes de 18 a 29 años, habría que eliminar del *marco* (el padrón de habitantes, el censo de población, el censo electoral, el de fecha más próxima) las unidades no incluidas en dicho intervalo de edad. Es preferible que la eliminación se haga antes de la extracción de la *muestra*.

e) *El marco ha de contener información suplementaria que ayude a localizar a las unidades elegidas.* Preferiblemente, su teléfono y dirección para facilitar su contactación.

Por ejemplo, en la encuesta por correo, la conveniencia de que los cuestionarios se manden personalizados (a una persona concreta), exige que el *marco* esté actualizado y que incluya los nombres y direcciones completas para que puedan enviarse los cuestionarios.

f) *El marco debe ser fácil de utilizar.* Aquellos que están en soporte magnético simplifican y automatizan bastante el proceso de extracción de la *muestra*. También reducen los costes del diseño de la *muestra* (en tiempo y dinero) y contribuyen a reducir *errores* en su selección.

El grado de cumplimiento de estos requisitos marcará los *errores de cobertura*. Se dan cuando el *marco muestral* “no cubre” a la *población* de estudio. Para Groves (1989: vi) es “el fracaso en dar a alguna persona de la población la oportunidad de selección para la muestra”. Los “no cubiertos” quedan excluidos del estudio y su efecto en la *calidad* de la investigación dependerá de la proporción que representen en el conjunto total de la población y de sus características diferenciales con los que sí han quedado cubiertos. Ésta ha sido la razón principal del mayor *error de cobertura* tradicionalmente atribuido a la encuesta telefónica. No tener teléfono confluía con un perfil sociodemográfico específico: un nivel educativo y de renta bajos, residir en hábitats rurales, edad más temprana del cabeza de familia y pertenecer a minorías étnicas (Couper y Hansen, 2002). Lo mismo es extensible, en la actualidad, a las encuestas web, por la aún lejana universalidad del uso de Internet (en el conjunto de la población).

El *error de cobertura* se conceptualiza como *error sistemático*. Representa una desviación de las *estimaciones muestrales*, respecto a los correspondientes *parámetros poblacionales*, que es “constante” en las diferentes repeticiones del estudio, si

se emplea el mismo *marco muestral* (las mismas unidades de la población quedan excluidas de la *muestra*). La magnitud del *error de cobertura* está determinada por:

Error de cobertura = (tasa de *no cobertura*) \times (Diferencia entre la media de Y o variable de interés en la *población cubierta* y la *no cubierta*)

$$E_{NC} = t_{NC} \times (\bar{Y}_C - \bar{Y}_{NC})$$

Donde: t_{NC} = proporción o tasa de *población no cubierta* en el *marco muestral*

\bar{Y}_C = media de Y (la variable que se analiza) en la *población cubierta*

\bar{Y}_{NC} = media de Y en la *no cubierta*

Un valor positivo expresa que el *marco* elegido supone una sobreestimación de la variable analizada; un valor negativo, una subestimación tanto en términos absolutos como relativos. Para facilitar la comparabilidad de variables en diferentes niveles de medición puede calcularse –siguiendo a Biemer y Lyberg (2003)– el *error de cobertura relativo*, que queda traducido en términos porcentuales:

$$\text{Error de no cobertura relativo} = t_{NC} \times \left(\frac{\bar{Y}_C - \bar{Y}_{NC}}{\bar{Y}_P} \right)$$

Donde \bar{Y}_P es la media de Y en la población de interés: $\bar{Y}_P = (t_C \times \bar{Y}_C) + (t_{NC} \times \bar{Y}_{NC})$

EJEMPLO DE CÁLCULO DEL ERROR DE COBERTURA

Una encuesta web realizada en España en 2012 concluye que la simpatía media de los españoles hacia los inmigrantes es de 6,8 (en una escala de 0 a 10). Una encuesta *cara a cara* simultánea obtiene que el 48% de la población aún carece de acceso a Internet. Entre los “no cubiertos” la simpatía media (\bar{Y}_{NC}) es 4,2. La encuesta web tendría un *error de cobertura* de:

$$E_{NC} = t_{NC} \times (\bar{Y}_C - \bar{Y}_{NC}) = 0,48 \times (6,8 - 4,2) = 1,248$$

Al ser un valor positivo significa que la encuesta web sobreestima la simpatía media hacia los inmigrantes en 1,248 puntos. En términos relativos el *error de cobertura* sería:

$$E_{NC_{relativo}} = t_{NC} \times \left(\frac{\bar{Y}_C - \bar{Y}_{NC}}{\bar{Y}_P} \right) = 0,48 \times \left(\frac{6,8 - 4,2}{5,552} \right) = 0,2248$$

$$\bar{Y}_P = (t_C \times \bar{Y}_C) + (t_{NC} \times \bar{Y}_{NC}) = (0,52 \times 6,8) + (0,48 \times 4,2) = 5,552$$

Un *error de cobertura relativo* del 22,48%. Una cuantía considerable debido al perfil sociodemográfico diferencial tan grande entre los “cubiertos” en la encuesta web y los excluidos, y en variables sociodemográficas claves en la explicación de la xenofobia. Principalmente su mayor edad y peor ubicación en las escalas educativas, ocupacionales y de ingresos, que definen su posición social (Cea D’Ancona y Valles, 2011).

Como el *error de cobertura* afecta a la “representatividad” de la *muestra*, conviene seguir la recomendación ya destacada por Fowler (1988: 22) de:

“Cuando se informen resultados, el investigador debe decir a los lectores a quién se dio y a quién no se dio una oportunidad de ser seleccionado, cómo aquellos omitidos diferían de los otros y si había o no algunas personas muestreadas para quienes las oportunidades de selección no eran conocidas”.

Las actuaciones ante el *error de cobertura* exigen tener información sobre el *marco muestral*: en qué medida cubre a la *población* de interés. La información disponible sobre la *población* ha de cruzarse con la proporcionada en el *marco*. Una vez comprobada la “calidad” del *marco*, el *error de cobertura* puede reducirse:

- a) *Restringiendo la población de estudio a la incluida en el marco muestral*. El *error de cobertura* desaparece cuando ambas poblaciones coinciden.

Por ejemplo, si un estudio sobre la situación laboral de los sociólogos se circunscribe a los incluidos en los directorios de los colegios de licenciados y doctores en ciencias políticas y sociología, limitar la población de estudio a sociólogos “colegiados” (dado que no todos los sociólogos están colegiados, a diferencia de otros profesionales superiores como abogados o médicos).

- b) *Conjugar varios marcos*, cuando se prevea que uno no cubre toda la *población* de interés.

Por ejemplo, no limitarse a los directorios de los colegios profesionales, sino conjugarlo con los listados de doctores y licenciados en las diferentes facultades de sociología en un período de tiempo concreto.

La utilización conjunta de varios *marcos* exige eliminar duplicidades, aparte de comprobar inclusiones erróneas. Tarea facilitada si se hallan en so-

porte magnético (“casar” datos por nombre, dirección, teléfono, DNI u otros datos proporcionados en el *marco*).

- c) *Identificar las “omisiones” y utilizarlas como variables de cuota* en el diseño de la muestra. Como indica Wert (2000: 208),

“El hecho de que en los hábitats más pequeños haya un equipamiento telefónico menor no lleva a subrepresentar tal hábitat, puesto que el hábitat suele ser un criterio de estratificación. El hecho de que las personas de más edad presenten una menor incidencia de teléfono en el hogar no comporta, en la mayoría de los casos, un sesgo en la muestra final puesto que la edad suele considerarse como variable de cuota”.

En referencia a la encuesta telefónica, Keeter (1995) propone el uso de “hogares transitorios”: aquellos que han experimentado un cambio en su situación de usuario de telefonía fija (acaban de incorporarse al directorio telefónico o se han dado de baja). El acceso a esta población permite, en parte, conocer las características de los hogares “sin teléfono”. De todas formas, cuando se quiera extraer información de la *población* “no cubierta” en encuestas telefónicas y web, *por ejemplo*, puede aplicarse otro método de encuesta (*cara a cara* e, inclusive, por correo). Incrementaría los costes de la investigación, aunque se amortizarían si la información relativa a los “no cubiertos” pudiese utilizarse en encuestas posteriores próximas en el tiempo.

- d) *Ajustes estadísticos mediante la ponderación*. Exige disponer de información complementaria “actualizada”, bien de otros estudios que incluyan a los “no cubiertos”, bien de datos censales. En la medida en que las variables utilizadas para la *ponderación* se hallen relacionadas con el problema de investigación y diferencien a la *población* cubierta de la “no cubierta”, se logrará mejorar las estimaciones muestrales. En caso contrario, no. Las variables que suelen utilizarse para *ponderar* son las presentes en los registros censales; preferiblemente, las relativas a estatus social (educación, ocupación, ingresos), edad, género, hábitat, etnia. Pero éstas no siempre están relacionadas con la variable de estudio.

El coeficiente de *ponderación* resulta de la estimación del peso de la *población* “cubierta” respecto del conjunto de la población, en la variable analizada. Se obtiene dividiendo el porcentaje de la *población* total entre el porcentaje de la *población* “cubierta”, en cada categoría de la variable elegida para la *ponderación*. Las estimaciones muestrales se multiplicarán por los *coeficientes de ponderación* respectivos. De la *ponderación* se habla con más detenimiento en los apartados 7.3.3 (en *muestras estratificadas no proporcionales*) y 7.5 (como estrategia para reducir el *error de no respuesta*). Si bien se trata de un ajuste estadístico que no logra eliminar *errores de cobertura*. Sólo compensa su efecto en las estimaciones muestrales.

7.2.3. Determinación del tamaño de la muestra

A la elección del *marco* sigue la determinación de cuántas unidades de la población incluir en la *muestra*. En esta decisión intervienen diferentes factores:

- a) *El tiempo y los recursos* disponibles para la realización del estudio.

Por ejemplo, en la encuesta telefónica la *muestra* puede ser más amplia (que en la encuesta *cara a cara*) y abarcar diferentes grupos de población, al conllevar un coste económico y temporal inferior. En encuestas a poblaciones específicas, como profesionales (profesores, sociólogos, médicos, jueces...) u organizaciones (hospitales, universidades, empresas...), el mayor esfuerzo para contactarlos y conseguir que participen en el estudio lleva a reducir el tamaño de la *muestra*. Más en el caso de la encuesta *cara a cara*; no tanto en la encuesta por correo, que suele obtener un mayor éxito (en porcentaje y calidad de las respuestas) en poblaciones de nivel educativo alto.

- b) La *modalidad de muestreo* elegida para alcanzar los objetivos del estudio. Cada tipo requiere unas exigencias específicas de tamaño y selección muestral. Los muestreos *no probabilísticos* (no enfocados a la inferencia estadística) precisan tamaños muestrales inferiores a los *probabilísticos* (cuya meta es el logro de una *muestra* “representativa”, que exige un mayor tamaño muestral). Dentro de los *muestreos probabilísticos*, el *aleatorio estratificado* requiere un mayor tamaño muestral. En los *no probabilísticos*, el muestreo *por cuotas*.
- c) La *diversidad de los análisis* previstos. Antes de diseñar la *muestra* conviene anticipar la variedad de análisis a realizar para alcanzar los objetivos de la investigación.

Por ejemplo, la correcta aplicación del análisis de *regresión lineal múltiple* exige al menos 40 casos por cada variable independiente introducida en el análisis (Cea D’Ancona, 2002a). En el análisis de *tablas de contingencia*, una regla es que el tamaño de la muestra sea al menos diez veces el número de celdillas de la tabla (Blaikie, 2010). En general, cuanto más se segmente la *muestra*, mayor tamaño muestral se precisa para que las estimaciones muestrales tengan significatividad estadística.

- d) La *varianza o heterogeneidad poblacional*. A mayor heterogeneidad (*varianza* poblacional), mayor tamaño de la *muestra* para que pueda reunir la diversidad de la *población*. Universos homogéneos precisan menos casos para representar a la *población*.

Todo conocimiento de la *varianza* poblacional permite reducir el tamaño de la *muestra* porque, cuando se desconoce, se aplica el supuesto de máxima heterogeneidad (P y $Q = 50 \times 50$) incrementando el tamaño muestral (cuadro

7.1): tomar el producto de las probabilidades “P” (de aparición de un suceso) y “Q” (no ocurrencia del suceso o evento: es igual a “1 – P”), equivalente a la *varianza* poblacional, en igual magnitud. Ambas probabilidades valdrían 0,50. Pero no siempre se conoce la *varianza* poblacional, a menos que se haga un estudio previo. Información que es factible en encuestas que se repiten con periodicidad (barómetros del CIS, eurobarómetros de la Unión Europea, la Encuesta de Población Activa del INE...) y cuando el *pretest* del cuestionario se realiza aplicando muestreo probabilístico (capítulo 6).

CUADRO 7.1. Tamaño muestral para poblaciones infinitas a un nivel de confianza del 95,5% (2σ)

<i>Límites de error (%) para $\pm 2\sigma$</i>	<i>Valores estimados de P y Q (%)</i>					
	1/99	10/90	20/80	30/70	40/60	50/50
0,1	39.600	360.000	640.000	840.000	960.000	1.000.000
0,5	1.584	14.400	25.600	33.600	38.400	40.000
1,0	396	3.600	6.400	8.400	9.600	10.000
1,5	176	1.600	2.844	3.733	4.267	4.444
2,0	99	900	1.600	2.100	2.400	2.500
2,5	63	576	1.024	1.344	1.536	1.600
3,0	44	400	711	933	1.067	1.111
3,5	32	294	522	686	784	816
4,0	25	225	400	525	600	625
5,0	16	144	256	336	384	400

Por ejemplo, para un margen de error de $\pm 3,5\%$, el tamaño de la *muestra* para una población de más de 100.000 unidades es de 294 unidades, si $P = 10$ y $Q = 90$. Si P y Q fuesen, respectivamente, 30 y 70, la *muestra* aumentaría más del doble: 686 unidades. Para el supuesto de máxima heterogeneidad (P y $Q = 50$), la *muestra* sería de 816 unidades (casi el triple del supuesto inicial de elevada homogeneidad: $P = 10$ y $Q = 90$).

- e) *El margen de error máximo admisible*: a mayor *error*, menor tamaño muestral (cuadro 7.1). Los incrementos de *tamaño muestral* repercuten en una mayor precisión de la estimación de los parámetros poblacionales, con la consiguiente reducción del *error muestral*.

La figura 7.1 representa la relación negativa entre el *error* y el *tamaño de la muestra*. Para su elaboración se han tomado los valores que aparecen en el cuadro 7.1, que corresponden a la situación más habitual: desconocimiento de la *varianza* poblacional ($P = Q = 50$). Como puede observarse, conforme

aumenta el tamaño de la *muestra*, descende el *error*. Por ejemplo, si la *muestra* se amplía de 816 a 2.500 unidades, el *error* disminuye de 3,5 a 2%. A partir de $\pm 2\%$ de *error* se disparan los aumentos en tamaño *muestral* para alcanzar una misma ganancia en reducción de *error* (con los consiguientes aumentos en costes de la investigación). En consecuencia, habrá que encontrar el punto medio entre *tamaño* y *error*. No son deseables *errores* superiores a $\pm 4\%$.

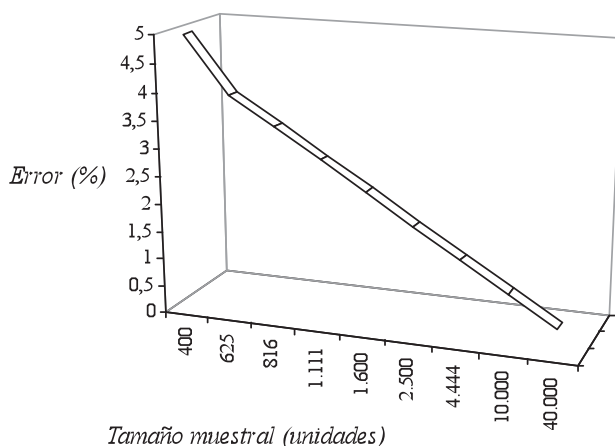


FIGURA 7.1. Variaciones error y tamaño muestral para $P = Q = 50$ y nivel de confianza del 95,5%.

- f) El nivel de confianza de la estimación de los parámetros poblacionales. A mayor nivel de *confianza* (o “probabilidad” de que las estimaciones muestrales se ajusten a la realidad), mayor *tamaño muestral* (cuadro 7.2). Los niveles de confianza corresponden a áreas bajo la curva normal acotadas por distintos valores de *desviación típica* (denominada *sigma* (σ) en referencia a la desviación poblacional o unidades Z). De ellos el más habitual es 2σ , que supone un 95,5% de probabilidad de acertar en la estimación. La plena confianza (exactamente un 99,7%) se alcanza con 3, que incrementa más del doble el tamaño de la muestra respecto del nivel de confianza habitual (95,5%).

Por ejemplo, para un error de $\pm 2\%$ y varianza $P = Q = 50$, el tamaño muestral es de 2.500 unidades, para un nivel de confianza de 95,5%. Si se incrementa a 99,7%, el tamaño de la *muestra* aumenta a 5.627 unidades. La ganancia en precisión es ínfima, mientras que los costes del estudio serían más del doble de los presupuestados para un nivel de confianza del 95,5%.

CUADRO 7.2. Tamaño muestral para poblaciones infinitas a un nivel de confianza del 95,5% (2σ) y 99,7% (3σ)

Límites de error (%)	Nivel de confianza (%)	Valores estimados de P y Q (%)				
		10/90	20/80	30/70	40/60	50/50
$\pm 1,0$	95,5	3.600	6.400	8.400	9.600	10.000
	99,7	8.100	14.400	18.900	21.600	22.500
$\pm 2,0$	95,5	900	1.600	2.100	2.400	2.500
	99,7	2.025	3.600	4.725	5.400	5.627
$\pm 2,5$	95,5	576	1.024	1.344	1.536	1.600
	99,7	1.296	2.304	3.024	3.456	3.600
$\pm 3,0$	95,5	400	711	933	1.067	1.111
	99,7	900	1.600	2.100	2.400	2.500
$\pm 4,0$	95,5	225	400	525	600	625
	99,7	506	900	1.181	1.350	1.406

Todos los aspectos referidos intervienen en el cálculo del *tamaño muestral*. La fórmula genérica para una *muestra aleatoria (simple o sistemática)* sería la siguiente, cuando el universo o población fuese *infinita* (integrada por más de 100.000 unidades):

$$n = \frac{Z^2 \hat{S}^2}{E^2} \quad \text{y} \quad n = \frac{Z^2 \hat{P}\hat{Q}}{E^2}$$

Donde: “ Z ” es el valor del área bajo la curva normal. Representa las unidades de *desviación típica* que corresponden al *nivel de confianza* elegido (2σ o 3σ): nivel de confianza de que las estimaciones se ajusten a la realidad.

“ \hat{S}^2 ” es la *varianza* poblacional estimada. Equivale al producto de las proporciones \hat{P} y \hat{Q} , siendo $\hat{Q} = 1 - \hat{P}$.

“ E ” denota el *error* máximo permitido que el investigador establece a priori. La precisión buscada en la estimación de los parámetros poblacionales.

Si el universo lo forman 100.000 unidades o menos, es una *población finita* y habría que introducir un factor de corrección en las fórmulas, quedando de la siguiente manera:

$$n = \frac{Z^2 \hat{S}^2 N}{E^2 (N-1) + Z^2 \hat{S}^2} \quad \text{y} \quad n = \frac{Z^2 \hat{P} \hat{Q} N}{E^2 (N-1) + Z^2 \hat{P} \hat{Q}}$$

Donde “N” es el tamaño de la población.

En *muestras aleatorias estratificadas y por conglomerados* se introducen otras variaciones que se comentarán en sus apartados correspondientes (7.3.3 y 7.3.4).

EJEMPLO DE CÁLCULO DEL TAMAÑO DE LA MUESTRA

Se quiere conocer la estimación de voto en los próximos comicios electorales. Calcular el tamaño de la muestra para los siguientes supuestos:

- a) Una encuesta de ámbito nacional para un *error* máximo de $\pm 2,5\%$ y careciendo de información previa sobre la *varianza* poblacional. Para un *nivel de confianza* del 95,5%, el *tamaño de la muestra* sería:

$$n = \frac{Z^2 \hat{P} \hat{Q}}{E^2} = \frac{2^2 \times 50 \times 50}{2,5^2} = 1.600 \text{ unidades}$$

Elevar el *nivel de confianza* al 99,7% incrementa el tamaño de la *muestra* en más del doble. Exactamente a 3.600 unidades.

$$n = \frac{Z^2 \hat{P} \hat{Q}}{E^2} = \frac{3^2 \times 50 \times 50}{2,5^2} = 3.600 \text{ unidades}$$

- b) Si se parte de la predicción (a partir de un sondeo efectuado con anterioridad) de una expectativa de voto para el PSOE del 38%.

$$n = \frac{Z^2 \hat{P} \hat{Q}}{E^2} = \frac{2^2 \times 38 \times 62}{2,5^2} = 1.508 \text{ unidades}$$

$$n = \frac{Z^2 \hat{P} \hat{Q}}{E^2} = \frac{3^2 \times 38 \times 62}{2,5^2} = 3.393 \text{ unidades}$$

- c) Si la encuesta se realiza en un municipio donde residen 86.724 personas de 18 y más años:

$$n = \frac{Z^2 \hat{P}\hat{Q}}{E^2(N-1) + Z^2 \hat{P}\hat{Q}} = \frac{2^2 \times 50 \times 50 \times 86.724}{(2,5^2 \times 86.723) + (2^2 \times 50 \times 50)} = 1.571 \text{ unidades}$$

$$n = \frac{Z^2 \hat{P}\hat{Q}}{E^2(N-1) + Z^2 \hat{P}\hat{Q}} = \frac{2^2 \times 38 \times 62 \times 86.724}{(2,5^2 \times 86.723) + (2^2 \times 38 \times 62)} = 1.482 \text{ unidades}$$

7.3. Selección de las unidades de la muestra: muestreos probabilísticos y no probabilísticos

En la *representatividad* de la *muestra* afecta su tamaño y el procedimiento seguido para seleccionar las unidades que la forman: si todos los miembros de la *población* tienen igual probabilidad de ser elegidos. El cuadro 7.3 reúne los diferentes procedimientos de selección agrupados en métodos de muestreo *probabilísticos* y *no probabilísticos*, dependiendo de si la selección de unidades se realiza aplicando algún procedimiento aleatorio o por criterios ajenos al azar. El “azar” garantiza que:

- Cada unidad de la *población* incluida en el *marco muestral* tenga igual probabilidad de ser elegida para formar la *muestra*.
- La elección de cada unidad sea independiente de las demás.
- El cálculo del *error muestral* y la consiguiente inferencia estadística: estimación de parámetros y pruebas de significatividad.

En los métodos de *muestreo no probabilísticos* la conveniencia, el juicio o determinismo humano, u otros criterios subjetivos, son los que intervienen en el proceso de selección. Ello da cabida a discrecionalidades por parte del equipo investigador, repercutiendo en:

- La desigual probabilidad de las unidades de la *población* para formar parte de la *muestra*.
- La dificultad para estimar el *error muestral*.
- La introducción de *sesgos* en el proceso de elección de la *muestra*, que pueden invalidar los hallazgos de la investigación.

Todo lo cual incide en la no adecuación de los *métodos no probabilísticos* en investigaciones dirigidas a la inferencia estadística. Lo que no invalida su uso en indagaciones *exploratorias* (estudios piloto, *pretest* del cuestionario...), en estudios *cualitativos* (más interesados en profundizar en la información que en su representatividad es-

estadística) y en investigaciones sobre población “*marginal*” (prostitutas, drogadictos, homosexuales, inmigrantes ilegales...), de las que no se dispone de un registro exhaustivo que abarque toda la población y facilite su localización.

Un mismo diseño muestral puede compaginar distintos métodos de muestreo en fases diferentes de su desarrollo. La elección del método dependerá de la confluencia de cuatro factores básicos:

1. El *presupuesto económico* de la investigación.
2. El período de *tiempo* programado para su ejecución.
3. La existencia de un *marco muestral* válido que haga factible la selección muestral aleatoria.
4. El grado de *precisión* que se quiera dar a la indagación.

CUADRO 7.3. Métodos de muestreo

A) *Métodos de muestreo probabilístico*

1. Muestreo aleatorio simple
2. Muestreo aleatorio sistemático
3. Muestreo aleatorio estratificado
 - Estratificación proporcional
 - Estratificación no proporcional: afijación simple, óptima y otras no proporcionales
4. Muestreo aleatorio por conglomerados: monoetápico, bietápico y polietápico
5. Muestreo de áreas y rutas aleatorias
6. Métodos de selección de personas en una misma vivienda
 - No selección: primera persona que contesta
 - Método de Kish de selección aleatoria.
 - Método de selección de cuota sistemática
 - Método del cumpleaños

B) *Métodos de muestreo no probabilístico*

1. Muestreo por cuotas
 2. Muestreo estratégico o de “juicio”
 3. Muestreos circunstanciales: de “voluntarios”, “bola de nieve”
-

7.3.1. Muestreo aleatorio simple

Es el arquetipo de muestreo, en referencia al cual se formulan los cálculos básicos del *tamaño y error muestral*. Como método *probabilístico* exige la existencia de un *marco muestral* que cumpla las condiciones antes expuestas. Ha de ser un marco nominativo, que proporcione información que ayude a contactar a las unidades ele-

gidas de manera aleatoria (nombre, dirección, número de teléfono...). A cada unidad de la población del *marco* ha de acompañar un número de *identificación* ordenado de manera consecutiva. Si no figura anexo, habrá que asignárselo, antes de proceder a la extracción de la *muestra*. Ésta se hará siguiendo algún procedimiento de selección aleatoria de números: una tabla de números aleatorios, un programa informático u otro procedimiento que garantice los dos requisitos básicos:

- a) Cada unidad de la *población* ha de tener igual probabilidad de participar en la *muestra*. Probabilidad determinada por la *fracción de muestreo* (el cociente entre el tamaño de la *muestra* y la *población* total en el *marco muestral*: n / N).
- b) La selección *aleatoria* de todas las unidades que componen la *muestra*. Se recomienda extraer un número mayor de unidades que las inicialmente prefijadas en el tamaño de la *muestra* para hacer frente a imprevistos durante el trabajo de campo (“sustituciones” de integrantes originales de la *muestra* que no puedan contactarse o que rehúsen participar en el estudio, principalmente).

La elección de las unidades de la *muestra* puede hacerse *sin reemplazamiento* (cada unidad puede ser elegida sólo una vez) o *con reemplazamiento* (participa en elecciones sucesivas). Lo habitual es que sea *lo primero: sin que se produzca reemplazamiento*. Si el *marco muestral* está en soporte magnético, se agiliza bastante el procedimiento de selección aleatoria de la *muestra*. En este caso, el programa informático diseñado al efecto realiza las tareas correspondientes a:

- a) Numerar consecutivamente a todos los integrantes del *marco muestral*.
- b) Generar la serie de *números aleatorios*.
- c) Seleccionar e imprimir la *lista* de las unidades finalmente elegidas para formar la *muestra* (aquellas a las que correspondan los números elegidos al azar).

En caso de tener que recurrir al procedimiento tradicional, las *tablas de números aleatorios* (que incluyen combinaciones de números extraídos al azar para que cada dígito tenga igual probabilidad de figurar en cualquier punto de la tabla), el modo de proceder sería el siguiente:

- Primero, se elige un *punto de partida* al azar: una columna o fila de la tabla. Puede comenzarse por cualquier lugar y moverse en cualquier dirección (hacia arriba, abajo, a la derecha o a la izquierda). La elección no ha de responder a un conocimiento previo de la distribución de los números de la tabla (introduciría *sesgos* en la *muestra*).
- El número de *dígitos* a seleccionar (en la tabla) ha de corresponder con el número de la *población* en el *marco muestral*.

Por ejemplo, si el *marco* estuviese compuesto por 58.420 unidades (5 dígitos), habría que numerar cada unidad de la población con 5 dígitos: desde el primero (00001) hasta el último (58.420), siguiendo un orden consecutivo. En la tabla se irían escogiendo números con 5 dígitos, desconsiderándose las cifras que superen la última unidad poblacional numerada en el listado (58.420).

- La unidad de la *población* en el *marco* al que corresponda el número extraído pasa a formar parte de la *muestra* (salvo que en el *marco* no figure información que permita su localización: dirección postal, teléfono...).

EJEMPLO DE MUESTREO ALEATORIO SIMPLE

En una universidad, que cuenta con 2.356 docentes, se quiere conocer su valoración de la reforma universitaria. Para un *error* de $\pm 2,5\%$ y un *nivel de confianza* de 95,5% la *muestra* estaría integrada por 953 docentes.

$$n = \frac{Z^2 \hat{P} \hat{Q} N}{E^2 (N-1) + Z^2 \hat{P} \hat{Q}} = \frac{2^2 \times 50 \times 50 \times 2.356}{(2,5^2 \times 2.355) + (2^2 \times 50 \times 50)} = 953$$

El listado de profesores está ordenado alfabéticamente y a cada docente se le asigna un número consecutivo. Se procede a la elección aleatoria de números mediante un programa informático. Si fuera aplicando una *tabla de números aleatorios*, los números elegibles han de ser inferiores o igual al tamaño de la *población*: 2.356. Han de seleccionarse al menos 953 números (pueden elegirse algunos más para imprevistos en el *trabajo de campo*).

A continuación se extraiga un fragmento de una *tabla de números aleatorios* (que figura al final del capítulo). Del extracto se deduce que la decisión tomada ha sido comenzar por la primera columna, escogiendo los cuatro primeros dígitos, continuar hacia abajo hasta el final de la columna, e ir alternando los cuatro primeros con los cuatro últimos dígitos. Se descartan los números superiores a 2.356. Los profesores elegidos serían a los que correspondan los siguientes *números de identificación*: 631, 1.472, 724, 1.726, 544, 2.341, 219..., así hasta completar al menos el tamaño muestral. La *probabilidad de selección* de cada unidad de la población en el *marco muestral* es 0,199, al ser la

$$\text{fracción de muestreo} = \frac{n}{N} = \frac{953}{2.356} = 0,404$$

CUADRO 7.4. Extracto de una tabla de números aleatorios

9675	4	1	7676	5565	9	4	4105	4736	1	3	4833
3435	7	8	6040	5336	4	7	1726	4569	0	6	6334
0631	8	3	7403	4992	7	5	7715	5042	3	6	7372
6211	1	5	2820	0724	3	7	9931	8929	2	8	4767
4753	4	0	9243	6787	9	0	0544	2341	0	1	2740
9861	4	7	5993	8446	0	6	2846	5984	4	1	4922
2485	6	0	3648	4489	8	0	9351	9879	5	1	8644
9688	7	1	2479	8062	1	6	6223	8608	5	7	8285
9080	1	2	1472	4281	5	7	7408	3739	0	7	6766
5516	5	7	7312	8366	6	3	6028	2842	0	7	0219

La exigencia de tener un *marco muestral* nominativo y numerado limita el uso del muestreo *aleatorio simple* en poblaciones de elevada magnitud.

Aun disponiendo de procedimientos informáticos que agilicen la selección de unidades muestrales, sería impracticable un listado, *por ejemplo*, de todos los hogares de un país; no cuando el estudio se circunscribe a un pequeño municipio.

El muestreo *aleatorio simple* se adecua a poblaciones pequeñas o circunscritas a un espacio delimitado. Pero no en aquellas espacialmente desperdigadas, salvo que el *trabajo de campo* no conlleve el desplazamiento de entrevistadores (caso de las encuestas *telefónicas* o por *correo*).

Enviar a un entrevistador a una barriada apartada o núcleo de población aislada para hacer una o dos entrevistas y, después, desplazarse a otro lugar para realizar otra, es impracticable por su elevado coste en tiempo (el número de entrevistas por día sería inferior) y dinero (desplazamiento, dietas). La mayor dispersión de la *muestra* alcanzada con este método de muestreo no supondría un coste añadido, si la encuesta es *telefónica* o por *correo*, y ganaría en representatividad.

El cuadro 7.5 resume las *ventajas* e *inconvenientes* principales del muestreo *aleatorio simple*.

CUADRO 7.5. Ventajas e inconvenientes del muestreo aleatorio simple

VENTAJAS	INCONVENIENTES
<ul style="list-style-type: none"> – Facilidad de los cálculos estadísticos – Elevada probabilidad de equiparabilidad entre la muestra y la población – Obtención de una muestra autoponderada (igual probabilidad de elección de cada unidad de la población en el marco muestral) 	<ul style="list-style-type: none"> – Requiere listar y numerar consecutivamente cada unidad del marco muestral – Un método de selección monótono y arduo en muestras y poblaciones de gran magnitud – La dispersión alcanzada en la muestra incrementa los costes de la investigación en estudios donde se desplazan investigadores o entrevistadores

7.3.2. Muestreo aleatorio sistemático

Para Sudman (1983: 170) es “muestreo aleatorio pseudosimple”, cuando el *marco muestral* esté aleatoriamente ordenado; para Lohr (2000: 43), es “una forma de muestreo por conglomerados”, aunque matizando “desde el punto de vista tecnológico”. Esta consideración responde a la imposibilidad de que dos unidades consecutivas del *marco* acaben siendo elegidas para componer la *muestra* (salvo que el *coeficiente de elevación* fuese “1”; lo cual es bastante improbable).

Al igual que el muestreo *aleatorio simple*, precisa un *marco muestral* nominativo, que incluya datos de identificación que posibiliten la contactación de las unidades de población que sean aleatoriamente seleccionadas. Pero no necesita que el *marco* sea un listado, ni que esté numerado. Pueden ser fichas de alumnos, papeletas, expedientes de pacientes, incluso la propia presencia física de unidades de la población (*por ejemplo*, personas que salen de una discoteca, bajan de un avión o entran en un centro comercial). Tampoco requiere la elección continua de números aleatorios. Se elige al azar sólo la primera unidad de la muestra. El resto, mediante la suma continua del *coeficiente de elevación*.

El *coeficiente de elevación* resulta del cociente entre los tamaños de la *población* y la *muestra* (N/n). El *marco muestral* se secciona en una serie agrupaciones o “conglomerados” igual al tamaño muestral; su amplitud (número de integrantes) la marca dicho coeficiente. Éste expresa el número de veces que la *muestra* se halla contenida en la *población*.

Una vez calculado, se elige al azar un número con la única condición de que sea inferior al *coeficiente de elevación*. A este número se suma el *coeficiente* de manera sucesiva hasta completar el tamaño de la muestra. Como en el muestreo *aleatorio simple*, se recomienda extraer un “excedente” de unidades muestrales para las *sustituciones*. Este excedente se consideraría antes de extraer la *muestra*, para el cálculo del *coeficiente de elevación*, y así evitar tener que “volver” al *marco muestral* en busca de sustitutos para quienes acaban no participando en la investigación. En caso contrario, la elección de “sustitutos” llevaría a repetir el proceso de selección hasta completar las unidades a sustituir: calculando un nuevo *coeficiente de elevación* y, a ser posible, considerando sólo aquellas unidades de la población que antes no fueron seleccionadas (*muestreo aleatorio sin reemplazamiento*). Éstas se dividirán entre el número total de unidades que aún se precisan para completar la *muestra*.

EJEMPLO DE MUESTREO ALEATORIO SISTEMÁTICO

Si la elección de profesores universitarios fuese mediante muestreo *aleatorio sistemático*, el procedimiento sería:

Primero se calcula el *intervalo de selección*, determinado por el *coeficiente de elevación* $\frac{N}{n} = \frac{2.356}{953} = 2,47$. El *marco* quedaría seccionado en 953 conglomerados de 2,47 unidades.

- Se elige al azar un número igual o inferior a “2” (redondeando). *Por ejemplo*, el “2”. La unidad de la población en el *marco* a la que corresponda dicho número constituirá la primera unidad elegida para formar la muestra.
- La extracción de las 952 unidades restantes se hará sumando “sistemáticamente” el *coeficiente de elevación* a los números elegidos, a partir del primer número extraído al azar. Al 2 seguiría el 4, 6, 8, 10, 12, 14, 16... (en este caso, todos los números pares) hasta completar las 953 unidades de la *muestra* inicial.

Para que la *muestra* se asemeje a la extraída aplicando muestreo *aleatorio simple*, es condición que las unidades de la población incluidas en el *marco muestral* estén colocadas de manera “aleatoria”; que su orden no responda a ningún criterio que pueda estar relacionado con el problema de investigación.

En el *ejemplo* anterior sería ordenar los profesores por orden alfabético (primer apellido), por los últimos dígitos de su DNI (Documento Nacional de Identidad) u otro criterio que garantizase su disposición aleatoria.

El cuadro 7.6 resume las *ventajas* y los *inconvenientes* principales del *muestreo aleatorio sistemático*:

CUADRO 7.6. Ventajas e inconvenientes del muestreo aleatorio sistemático

VENTAJAS	INCONVENIENTES
<ul style="list-style-type: none"> – Fácil aplicación. No precisa de la elección continua de números aleatorios – El <i>marco muestral</i> no se restringe a un listado (fichas, papeletas, expedientes, inclusive personas u otras unidades de población) 	<ul style="list-style-type: none"> – Precisa del recuento constante de las unidades de la población – Antes de la selección muestral, hay que “desordenar” el <i>marco muestral</i> (no ha de responder a ningún criterio que favorezca la elección de unidades concretas de la población)

7.3.3. Muestreo aleatorio estratificado

Un método muy aplicado cuando se dispone de información que permita clasificar a la población de estudio. Ésta quedará clasificada en grupos o “estratos” mutuamente excluyentes (internamente homogéneos y diferentes de los otros grupos), de-

finidos por la característica que se elija de las presentes en el *marco* muestral. Ello posibilita:

- a) *Garantizar la presencia en la muestra de grupos de población* que, por su menor peso en la población, tendrían escasa probabilidad de ser seleccionados para componer la *muestra* aplicando otros procedimientos aleatorios (simple, sistemático u otros), siendo su estudio de interés.

Por ejemplo, clasificar a los estudiantes por curso académico garantiza que alumnos de menor presencia (caso de alumnos de doctorado) tengan representación en la *muestra*. Al igual que clasificar a los profesores por facultades o departamentos; o a los municipios por tamaño de hábitat.

- b) *Aplicar métodos de selección aleatorios diferentes en cada estrato*, dependiendo de la dificultad para contactar con las unidades de la *muestra*.
- c) *Mayor precisión de las estimaciones de los parámetros poblacionales* en los diferentes grupos o estratos, al reducirse su variabilidad. Pero sólo cuando los estratos son internamente homogéneos y diferentes entre sí, y en variables relacionadas con el problema de investigación. De utilizarse más de una variable de *estratificación*, conviene que sean variables poco correlacionadas entre sí para conseguir grupos heterogéneos unos respecto de otros.

La elección de las variables de *estratificación* está condicionada a las presentes en el *marco muestral*. *Por ejemplo*, en estudios de ámbito nacional e internacional, primero se procede a la *estratificación* por ubicación geográfica: *ámbito territorial* (país, comunidad autónoma, provincia, municipio), *tipo de hábitat* (urbano, semiurbano, rural) o *tamaño de hábitat* (número de habitantes). Después, se realizan otras *estratificaciones*, en función de los objetivos del estudio y la información proporcionada por el *marco muestral*. Cuando la *estratificación* comprende diferentes fases, el orden de las variables dependerá de su relevancia en la explicación del problema de investigación. Como la primera variable de *estratificación* es la que más discrimina, conviene que se elija la de mayor relevancia para la primera fase; en la segunda fase, la segunda más importante, y así consecutivamente. Las variables elegidas han de ser categóricas. En caso contrario, se procederá a su categorización.

Tras la clasificación de la población en estratos, se procede a la *afijación* o distribución de la *muestra* en los diferentes estratos. Existen diferentes procedimientos de *afijación*. Éstos caben agruparse dentro de dos amplias modalidades de *estratificación*: *proporcional* y *no proporcional*.

- a) *Estratificación proporcional*: la *afijación* es “proporcional” al peso del *estrato* en el conjunto de la población. A los estratos que aglutinen un mayor número de unidades de población les corresponderá un mayor tamaño muestral; los de menor peso tendrán una menor representación en la *muestra*. Es el

criterio de *afijación* más aplicado en la investigación social. Se multiplica la proporción que representa el estrato en la *población* por el tamaño de la *muestra* a afijar.

- b) *Estratificación no proporcional*: la *afijación* de la *muestra* no es proporcional a su peso en el conjunto de la población. La desproporcionalidad puede responder a motivos diferentes: que todos los estratos tengan igual presencia en la *muestra* para posibilitar la realización de análisis comparativos entre ellos (*afijación simple*); considerar la varianza del *estrato*, además de su peso en el conjunto de la población (*afijación óptima*); o introducir desproporcionalidad afijando un mayor número de unidades muestrales a estratos de poco peso poblacional o a la inversa (*afijación no proporcional*):

- *Afijación simple*: el tamaño de la *muestra* se distribuye en igual magnitud entre los diferentes *estratos* (indistintamente de su peso en el conjunto de la población). Se quiere garantizar que todos tengan igual base muestral, que facilite la comparación entre ellos y la obtención de estimaciones muestrales igualmente significativas. El inconveniente es que favorece a los estratos de menor peso poblacional en menoscabo de aquellos de mayor magnitud. Se suma la exigencia de *ponderar* la *muestra* (devolverle su debida proporcionalidad) para el análisis conjunto de los estratos.

La *afijación simple* se obtiene dividiendo el tamaño de la *muestra* entre el número de *estratos*.

- *Afijación óptima*: la *muestra* se afija considerando el peso del estrato y su varianza en la variable de estratificación. A los estratos de mayor peso y heterogeneidad les corresponderá un mayor número de unidades muestrales; lo contrario a aquellos más homogéneos y de menor peso. Su aplicación requiere conocer la distribución de la varianza de la variable de estratificación en cada *estrato*. Información no disponible a menos que se realice un estudio previo de la población. Ello limita su uso, pese a su calificación “óptima” (más cuando los estratos presentan varianzas divergentes; en caso contrario, se obtendría una distribución muestral similar a la *afijación proporcional*).

Para su obtención se multiplica la *proporción* (o porcentaje) que representa el estrato en el conjunto de la población por su *varianza* correspondiente; se suman los productos y se calcula la proporción que cada uno representa respecto de la suma total. Esta proporción es la que se multiplica por la *muestra* a afijar, obteniendo el número de unidades muestrales en cada estrato.

- *Afijación no proporcional* “deliberada”: la desproporcionalidad de la *muestra* responde al criterio del investigador, que determina el número de unidades muestrales en uno o varios estratos. En los demás estratos, la *muestra* se afija de manera proporcional (excluyendo los estratos ya asignados). Aunque esta última afijación sea “proporcional”, la distribución

de la *muestra* total en el conjunto de los estratos no es proporcional. El análisis conjunto de los *estratos* exige, en consecuencia, la previa *ponderación* de la *muestra* en cada estrato.

EJEMPLO DE MUESTREO ALEATORIO ESTRATIFICADO

Se quiere evaluar la calidad de la enseñanza que reciben los estudiantes de la Universidad Complutense. En el presente curso hay un total de 74.096 alumnos matriculados en sus diferentes centros. Para garantizar que en la *muestra* estén representados alumnos de distintas enseñanzas, se *estratifica* atendiendo a dicho criterio. Un estudio previo concluye que el 64% de los estudiantes está satisfecho. El porcentaje varía entre el 45% de los alumnos de diplomaturas y el 72% de doctorados. Para un *error* máximo de 2,5% y un *nivel de confianza* del 95,5%, se obtiene un *tamaño muestral* de 1.446 alumnos a entrevistar. Estos se distribuyen entre los *estratos* atendiendo a los tres criterios principales de *afijación* (simple, proporcional y óptima):

$$n = \frac{Z^2 \hat{P} \hat{Q} N}{E^2 (N-1) + Z^2 \hat{P} \hat{Q}} = \frac{2^2 \times 64 \times 36 \times 74.096}{(2,5^2 \times 74.095) + (2^2 \times 64 \times 36)} = 1.446 \text{ unidades}$$

CUADRO 7.7. Distribución de la muestra en estratos

Enseñanza	Población	Varianza	Afijación		
			Simple	Proporcional	Óptima
Grados	22.073	2.475	289	431	470
Licenciaturas	33.220	2.139	289	648	610
Diplomaturas	6.681	2.496	289	130	143
Másteres	4.509	2.331	289	88	91
Doctorado	7.613	2.016	289	149	132
Total	74.096	2.304	1.445	1.446	1.446

Afijación simple: $\frac{1.446}{5} = 289,2$ unidades en cada estrato

Afijación proporcional: 0,298 x 1.446 = 430,91
 0,448 x 1.446 = 647,81
 0,090 x 1.446 = 130,14
 0,061 x 1.446 = 88,21
 0,103 x 1.446 = 148,94

Al considerar la proporción que representa cada estrato en el conjunto de la población, los estudiantes de licenciatura tendrán mayor presencia en la *muestra* (648); los de doctorado, la menor (88).

Afijación óptima:

CUADRO 7.8. Cálculo de la afijación óptima

(A) Porcentaje población	(B) Varianza	(C) (A) x (B)	(D) Proporción (C)	(E) Afijación óptima (D) x 1.446
29,8	2.475	73.755,0	0,325	469,95
44,8	2.139	95.827,2	0,422	610,21
9,0	2.496	22.464,0	0,099	143,15
6,1	2.331	14.219,1	0,063	91,10
10,3	2.016	20.764,8	0,091	131,59
100,0	2.304	227.030,1	1,000	1.446

En la *afijación óptima* los alumnos de licenciaturas de nuevo acaparan la mayor representación muestral (610), aunque en menor magnitud que en la *afijación proporcional* (648). Se debe a que es el segundo estrato más homogéneo o de menor varianza (2.139). En cambio, los alumnos de másteres oficiales son los de menor presencia (91), aunque ésta es ligeramente superior a la *afijación proporcional* (88). Se explica por su varianza (2.331), la tercera más elevada.

Cualquier procedimiento de *afijación no proporcional* requiere el cálculo de *coeficientes de ponderación*. Ha de devolverse a la *muestra* su debida proporcionalidad, antes de proceder a la obtención de estimaciones muestrales del conjunto de los estratos (no para análisis segmentados de cada estrato). Para ello se divide el porcentaje o proporción que representa el *estrato* en la población entre el que presenta en la *muestra*. Los “pesos” o coeficientes que resultan se multiplicarán por las estimaciones muestrales en cada estrato. A partir de lo cual se obtienen las estimaciones muestrales conjuntas, sumando los productos de los *coeficientes de ponderación* por las *estimaciones muestrales* en cada estrato, y dividiendo la suma total de los productos por el tamaño total de la *muestra*. En la generalidad de los paquetes estadísticos, la *ponderación* se cumplimenta de manera automática antes de la tabulación conjunta y posterior análisis de la información recabada. En la muestra “real”, no en la ponderada, será en la que después se calculen los *errores de muestreo* y la consiguiente precisión de las estimaciones muestrales (apartado 7.4).

EJEMPLO DE PONDERACIÓN EN UN MUESTREO ESTRATIFICADO NO PROPORCIONAL

Veamos cómo se calcularían los *coeficientes de ponderación* en el ejemplo anterior. Para ello se plantean dos supuestos: uno, para la *afijación simple*; dos, para la *estratificación no proporcional*, en la que se afija una porción determinada de la *muestra* total en un estrato concreto (la *muestra* restante se asignaría de forma proporcional entre los demás estratos). *Por ejemplo*, se incrementa la presencia de alumnos de másteres para que haya una base muestral suficiente que contribuya a la significatividad estadística de sus estimaciones: entrevistar a 120 alumnos en vez de 88 (que correspondería mediante *afijación proporcional*). Ese incremento en 32 alumnos de másteres supone una reducción en la representación de los estudiantes de otros cursos. Las 1.326 unidades muestrales que quedan por afijar ($1.446 - 120$) se distribuirán de manera proporcional entre los cuatro estratos restantes. Del reparto se excluye a los alumnos de másteres. Por lo que la población total será ahora de 69.587 estudiantes. Sobre esta base se calculan los porcentajes que representan cada estrato en la población (por ejemplo, $22.073 / 69.587 = 0,317$, multiplicado por la muestra por afijar, $1.326 = 420,6$; al igual se procede en los estratos restantes). Aunque esta última afijación sea proporcional, al sumarse los alumnos de máster (120), se está ante una *estratificación no proporcional*, con la exigencia correspondiente de calcular los coeficientes de *ponderación* como ilustra el cuadro 7.9:

CUADRO 7.9. Cálculo de los coeficientes de ponderación en una muestra estratificada no proporcional

Enseñanza	Población	(A) Porcentaje población	Afijación proporcional	Afijación no proporcional	(B) Porcentaje muestra no proporcional	Coeficiente ponderación (A/B)
Grados	22.073	29,8	431	421	29,1	1,024
Licenciaturas	33.220	44,8	648	633	43,8	1,023
Diplomaturas	6.681	9,0	130	127	8,8	1,023
Másteres	4.509	6,1	88	120	8,3	0,735
Doctorado	7.613	10,3	149	145	10,0	1,030
Total	74.096	100,0	1.446	1.446	100,0	

El *coeficiente de ponderación* se obtiene dividiendo el porcentaje de la *población* entre el porcentaje de la *muestra* en cada estrato (*por ejemplo*, $29,8/29,1 = 1,024$). Puede asimismo obtenerse dividiendo la *afijación proporcional* entre la *no proporcional* (*por ejemplo*, $431/421 = 1,024$). En ambos casos expresan cómo se desvía la *muestra* de su distribución proporcional a su peso en el conjunto de la *población*. Un coeficiente igual a “1”

significa que el tamaño muestral coincide con el afijado de forma proporcional; un coeficiente mayor de “1”, que se ha asignado un tamaño muestral inferior al proporcional (tanto más cuanto más se aleje de “1”); y un coeficiente inferior a “1”, que se han seleccionado más unidades muestrales de las obtenidas mediante afijación proporcional. En este ejemplo sólo ocurre en el estrato de estudiantes de másteres (0,735).

Para el supuesto de *afijación simple*, se procedería de igual forma. En este caso es en el estrato correspondiente a los estudiantes de licenciaturas (289) donde ha habido mayor distancia respecto a la muestra proporcional (648). Por lo que se obtiene un coeficiente de 2,242. Coeficiente por el que se multiplicarían los datos obtenidos en este estrato para el análisis conjunto de los alumnos de la Universidad Complutense (la respuesta de cada encuestado se multiplicaría por 2,24, mientras que las correspondientes a los estudiantes de másteres o doctorado lo serían por 0,304 y 0,516, respectivamente).

CUADRO 7.10. Cálculo de los coeficientes de ponderación para afijación simple

<i>Enseñanza</i>	<i>Población</i>	<i>Porcentaje población</i>	<i>Afijación proporcional</i>	<i>Afijación simple</i>	<i>Porcentaje muestra simple</i>	<i>Coeficiente ponderación</i>
Grados	22.073	29,8	431	289	20,0	1,491
Licenciaturas	33.220	44,8	648	289	20,0	2,242
Diplomaturas	6.681	9,0	130	289	20,0	0,450
Másteres	4.509	6,1	88	289	20,0	0,304
Doctorado	7.613	10,3	149	289	20,0	0,516
Total	74.096	100,0	1.446	1.445	100,0	

Los *coeficientes de ponderación* se introducen en el paquete estadístico al uso para que, previo al cálculo de las estimaciones muestrales en el conjunto de la población, se sumen los distintos productos de los coeficientes de ponderación por las estimaciones muestrales en cada estrato y, después, se divida por el tamaño total de la *muestra*. Esto posibilita la obtención de estimaciones “equilibradas” para el conjunto de la *muestra*.

Tras la afijación de la *muestra* procede la elección del procedimiento aleatorio a aplicar para la selección de las unidades muestrales en cada estrato. A diferencia del *muestreo por cuotas*, en el *estratificado* la selección de las unidades finales de la *muestra* se hace al azar (mediante muestreo aleatorio simple, sistemático, rutas aleatorias...), y no mediante criterios subjetivos (apartado 7.3.7). Las ventajas e inconvenientes del *muestreo aleatorio estratificado* se resumen en el cuadro 7.11.

CUADRO 7.11. Ventajas e inconvenientes del muestreo aleatorio estratificado

VENTAJAS	INCONVENIENTES
<ul style="list-style-type: none"> – Estimaciones muestrales más precisas, al haber menor <i>error</i> muestral – Garantiza la presencia en la <i>muestra</i> de los diferentes grupos de población en variables claves para la investigación – Pueden utilizarse distintos <i>marcos</i> muestrales y métodos de muestreo en cada estrato – Facilita la organización del <i>trabajo de campo</i> 	<ul style="list-style-type: none"> – Requiere más información del <i>marco</i> muestral que el muestreo aleatorio simple (para identificar la población de cada estrato); lo que puede resultar gravoso – Supone cálculos estadísticos de cierta complejidad

7.3.4. Muestreo aleatorio por conglomerados

De gran aplicación cuando la investigación abarca poblaciones dispersas y precisa del desplazamiento de miembros del equipo investigador. Al igual que el muestreo aleatorio *estratificado*, el de *conglomerados* procede a la selección aleatoria de conjuntos de población, que se denominan “conglomerados” (no “estratos”). Los *conglomerados* pueden ser demarcaciones territoriales (país, comunidad autónoma, municipio, distrito, áreas censales, viviendas...), organizaciones, instituciones o centros (colegios, hospitales, tribunales, centros penitenciarios...), inclusive conglomerados artificiales (como las urnas electorales). Se selecciona una muestra de *conglomerados* al azar y, en los elegidos, se extraen las unidades que compondrán la *muestra*, aplicando igualmente algún método de selección aleatoria. El cuadro 7.12 resume las diferencias principales entre el muestreo por *conglomerados* y el *estratificado*.

CUADRO 7.12. Diferencias entre el muestreo aleatorio estratificado y de conglomerados

ESTRATIFICADO	POR CONGLOMERADOS
La unidad de muestreo es <i>unitaria</i> (el individuo)	La unidad de muestreo es el <i>conglomerado</i> (o conjunto de individuos)
Se extrae una <i>muestra</i> en todos los grupos o <i>estratos</i> diferenciados de acuerdo con una o varias variables de estratificación	Se extrae una <i>muestra</i> aleatoria de <i>conglomerados</i> . Los integrantes de los conglomerados elegidos forman la <i>muestra</i>
Se busca la <i>homogeneidad</i> dentro del <i>estrato</i> y la heterogeneidad entre los <i>estratos</i> . Los grupos se forman para crear grupos internamente homogéneos respecto de la(s) variable(s) de interés. Reduce el <i>error</i> muestral	Se busca la <i>heterogeneidad</i> dentro del <i>conglomerado</i> . El <i>error muestral</i> disminuye, a medida que aumenta la heterogeneidad dentro del <i>conglomerado</i> . Pero no siempre es posible, por lo que presenta un mayor <i>error</i> muestral

El muestreo por *conglomerados* abarata los costes del trabajo de campo (reduciendo los desplazamientos de los miembros del equipo investigador).

Por ejemplo, si en un bloque de viviendas (unidad de muestreo *primaria* o *conglomerado*) se entrevista a 10 familias (unidad de muestreo *secundaria*), en vez de una sola familia, y las nueve restantes se seleccionan de otros bloques de viviendas (al aplicarse otra variedad de muestreo aleatorio), se ahorra en tiempo y dinero en la ejecución del *trabajo de campo*. En su contra, el incremento del *error muestral*: es más probable que las 10 familias aleatoriamente extraídas de un mismo bloque sean más homogéneas que las procedentes de 10 bloques diferentes.

Al ahorro en costes económico-temporales, se suma la dificultad de elaborar una lista exhaustiva de todos los integrantes de la población (para aplicar otro método aleatorio) como las dos razones principales de la aplicación del muestreo por *conglomerados*. Siguiendo las recomendaciones de Sudman (1976):

- a) Los *conglomerados* han de estar *bien definidos* y *delimitados*. Cada unidad de la población sólo puede pertenecer a un único conglomerado.
- b) El *número de elementos* que componen el *conglomerado* ha de ser *conocido* antes de su selección, aunque sea de manera aproximada.
- c) Los *conglomerados* elegidos han de ser *pocos*, si realmente se quiere reducir los costes de la investigación. Si se busca disminuir el *error muestral*, habrá que aumentar el número de conglomerados, reduciendo el número de unidades de la muestra a seleccionar en cada uno de ellos.
- d) Los *conglomerados* no tienen por qué estar idénticamente definidos en todos los lugares. Cuando se muestrean individuos u hogares en áreas urbanas o semiurbanas, los *conglomerados* suelen ser bloques o un conjunto de bloques. En cambio, en áreas rurales, los conglomerados suelen ser segmentos geográficos limitados por carreteras y fronteras naturales (como ríos y lagos).

El muestreo por *conglomerados* puede ser singular o *monoetápico* (si todas las unidades de los conglomerados elegidos componen la *muestra*), *bietápico* o *polietápico*. *Bietápico*, cuando la selección muestral prosigue dentro de cada conglomerado, en una segunda fase: en cada conglomerado elegido al azar se procede a la selección aleatoria de las unidades muestrales. Si el proceso no concluye aquí, se está ante un muestreo *polietápico* de tres, cuatro e, inclusive, más fases. Como señalan Lynn y Lievesley (1991), lo usual es que el muestreo *polietápico* sea de tres o cuatro fases.

Por ejemplo, una encuesta a médicos de familia. La *primera fase* consistiría en la selección aleatoria de áreas sanitarias. La *segunda fase*, la selección aleatoria de centros de salud dentro de las áreas que hayan sido seleccionadas. Cuando se entrevista a todos los médicos de los centros elegidos, se está ante un *muestreo de conglomerados bietápico*. Si se procede a la selección aleatoria de médicos, en cada centro de salud elegido, la *muestra* será *trietápica*. Los centros de salud constituyen los *conglomerados* y los médicos las *unidades de observación*.

EJEMPLO DE MUESTREO ALEATORIO POR CONGLOMERADOS

- A) Una encuesta *cara a cara* a 1.600 profesores universitarios para conocer su opinión sobre la reforma universitaria. Para reducir costes, se procede a un muestreo *por conglomerados polietápico*, que incluye las actuaciones siguientes:
- *Listar todas las universidades españolas* (públicas y privadas). Este listado constituye el *marco* inicial para la extracción de la *muestra*. Se estratifica por *tipo de universidad* (pública y privada) y *ubicación geográfica* (comunidad autónoma) para asegurar la representación de las unidades de muestreo *primarias* (universidades) según las variables de estratificación elegidas: dispersión geográfica y tipo de universidad. Interesa que las *unidades primarias de muestreo* sean bastante heterogéneas para alcanzar una mayor precisión. De acuerdo con Lynn y Lievesley (1991), no más de un 7% de la muestra total ha de proceder de una unidad primaria de muestreo
 - *Elegir una muestra de facultades* (de las universidades seleccionadas) mediante alguno de los procedimientos de selección *aleatoria* (*simple, sistemática, estratificada*).
 - Para cada facultad, *extraer una muestra de profesores aleatoria*, a partir de un nuevo *marco* muestral: el listado de los profesores de las facultades elegidas. Éstos podrían clasificarse por categoría profesional (catedrático, titular, contratado doctor, ayudante, asociado), dedicación (tiempo completo, parcial), y antigüedad docente, aplicando un *muestreo aleatorio estratificado o por cuotas*. Si se opta por el *estratificado*, la elección de los profesores será totalmente *aleatoria* (muestreo *simple, sistemático*); en el de *cuotas*, la elección será arbitraria, con la única condición de que se ajusten a las *cuotas* marcadas. También podría extraerse una muestra aleatoria de departamentos y entrevistar a todos los profesores que lo formen. Esto último supondría reducir el número de departamentos elegidos. O, podría optarse por una selección aleatoria de profesores en cada departamento. Lo cual permitiría ampliar el número de departamentos en la *muestra* y conseguir que ésta fuese más heterogénea, con la consiguiente reducción en *error muestral*. Es más probable encontrar más heterogeneidad entre profesores de departamentos y de facultades diferentes que entre aquellos de un mismo departamento y facultad.

La opción está entre aumentar el número de facultades, disminuyendo el número de profesores a entrevistar en cada una de ellas, o proceder a la inversa (reducir el número de facultades, incrementando el número de profesores a encuestar). La primera opción eleva los costes de la investigación, pero reduce el *error muestral*. Diferentes opciones posibles de selección de los 1.600 docentes se resumen en el cuadro 7.13.

CUADRO 7.13. Opciones elección número de facultades y profesores

NÚMERO DE FACULTADES	NÚMERO DE PROFESORES
5	320
10	160
20	80
30	53
40	40
50	32

Elegir 5 facultades (entrevistando a 320 profesores en cada una) es la opción más económica, pero de mayor *error muestral*. En el extremo contrario, la elección de 50 facultades (y de 32 docentes en cada una de ellas). Cuando los conglomerados son muy homogéneos, se recomienda ampliar el número de conglomerados en busca de heterogeneidad, que reduzca el *error*. Si fuesen heterogéneos, podría reducirse el número de conglomerados (aumentando las unidades a observar en cada uno de ellos). Es la *varianza del conglomerado* la que ha de determinar el número de conglomerados a elegir.

B) Muestreo por conglomerado desarrollado por el Servicio de Evaluación del CIDE: Mariano Álvaro Page (1989: "La técnica de muestreo: ejemplo práctico de su aplicación en las evaluaciones educativas". *REIS*, n.º 46: 173-182).

El estudio para evaluar la reforma del segundo ciclo de las enseñanzas medias incluye a los alumnos de bachillerato de todo el Estado español que estén cursando COU y segundo curso del bachillerato experimental; y a los alumnos de centros de formación profesional (en las ramas administrativa y comercial, electricidad y electrónica, metal). En el estudio para evaluar la Reforma del ciclo superior de la educación general básica, la población está constituida por los alumnos de 8.º curso de EGB, que cursan estudios en centros del Ministerio de Educación y Cultura (MEC). No se consideran (por razones económicas) ni a los alumnos de BUP y EGB que estudian en centros con menos de 30 alumnos en los cursos evaluados (COU, 2.º de Bachillerato Experimental y 8.º de EGB), ni a los alumnos de FP de los centros con menos de 20 alumnos en los cursos considerados.

Cada centro es un conglomerado y en cada uno se elige un número fijo de alumnos: 20 o 30. El tamaño de la *muestra* en un muestreo por *conglomerados* depende del *efecto del diseño*, además del tamaño de la muestra en un muestreo *aleatorio simple* (Kish, 1965; Azorin y Sánchez-Crespo, 1986). El *efecto de diseño* es la proporción de la varianza del muestreo aleatorio *simple* explicada por la varianza del muestreo por *conglomerados*:

$$ED = \frac{S_{MC}^2}{S_{MAS}^2}$$

donde " S_{MC}^2 " es la varianza del muestreo por conglomerados y " S_{MAS}^2 ", la varianza del muestreo aleatorio simple.

El efecto de diseño también se expresa así:

$$ED = 1 + (b - 1)\rho, \text{ donde "b" es el tamaño del conglomerado y } \rho = \frac{bS_c^2 - S_r^2}{(b - 1)S_r^2}$$

Es decir, el coeficiente " ρ " depende de la varianza de las puntuaciones medias de las escuelas, de la varianza total o varianza del conjunto global de elementos (estudiantes en este caso) y del tamaño de los conglomerados, que es fijo (30 o 20, según los casos en los trabajos del servicio de evaluación). A mayor varianza entre los centros, mayor coeficiente ρ . La estimación del coeficiente ρ se hace a partir de los datos recogidos en un estudio piloto o en aplicaciones anteriores a poblaciones similares a las del estudio en cuestión.

El tamaño de la muestra por conglomerados N_{MC} depende del tamaño de la muestra del muestreo aleatorio simple (N_{MAS}) y del efecto de diseño:

En el caso de los estudios del servicio de evaluación del CIDE, que trabaja con un nivel de confianza del 95,5% y un margen de error del 5%, y que tienen coeficientes $N_{MC} = N_{MAS} [1 + \rho(b - 1)]$ entre .2 y .3, las muestras serían:

- COU: $N_{MC} = 400 [1 + .25 (30 - 1)] = 3.300$
- FP: $N_{MC} = 400 [1 + .25 (20 - 1)] = 2.300$
- EGB: $N_{MC} = 400 [1 + .25 (30 - 1)] = 3.300$

Los N_s mínimos de las muestras serían:

- COU: 3.300 sujetos y 110 conglomerados
- FP: 2.300 sujetos y 115 conglomerados
- EGB: 3.300 sujetos y 110 conglomerados

Las muestras "reales" se aumentaron en COU (6.000 sujetos y 200 centros) y FP (3.000 sujetos y 150 centros). Los muestreos realizados por el servicio de evaluación son *bietápicos*. En la primera fase la unidad muestral es el centro; en la segunda, el alumno.

En la primera fase se utiliza un muestreo aleatorio *estratificado con afijación proporcional y por conglomerados*. En esta fase se empieza definiendo qué son "unidades primarias de muestreo", que están delimitadas por las transferencias o no de las competencias educativas. Son territorio MEC: Andalucía, Canarias, Cataluña, Comunidad Valenciana, Euskadi y Galicia. Dentro de cada *unidad primaria* se hacen muestreos aleatorios *estratificados con afijación proporcional*. Las variables para la estratificación son el tamaño del *municipio* (menos de 10.000 habitantes, 10.001-50.000, 50.001-500.000, más de 500.000, Madrid y Barcelona) y *tipo de centro* (públicos y privados). En cada Comunidad Autónoma la asignación proporcional por estratos, según hábitat y tipo de centro, queda como aparece en el cuadro 7.14:

CUADRO 7.14. Estratificación proporcional por hábitat y centro

	PÚBLICOS			PRIVADOS		
	<i>Población</i>	<i>Muestra alumnos</i>	<i>Muestra centros</i>	<i>Población</i>	<i>Muestra alumnos</i>	<i>Muestra centros</i>
Menos de 10.000	7.572	180	6	1.603	30	1
10.001 – 50.000	13.408	330	11	2.220	30	1
50.001 – 500.000	38.003	960	32	12.841	240	8
Más de 500.000	3.533	90	3	2.322	60	2
Madrid	19.598	480	16	18.024	360	12
Totales	82.114	2.040	68	37.010	720	24

Después se eligen los centros dependiendo de la probabilidad de elección ($p_i = n_i / N_{Ei}$; donde “ N_{Ei} ” es el tamaño del estrato correspondiente). Primero, se averigua el número de alumnos que tiene cada centro del curso a evaluar (tamaño del centro). Después se realiza una acumulación de frecuencias por centro.

Por ejemplo, supongamos que hay tres centros con 100, 75 y 150 alumnos, respectivamente. El centro 1 tendría asignados los números comprendidos entre 1 y 100, el centro 2 los comprendidos entre 101 y 175, y el centro 3 entre 176 y 325.

Por último, se generan números aleatorios. El centro que incluya, entre los números asignados en el paso anterior, el número generado aleatoriamente, pasa a formar parte de la muestra.

En encuestas nacionales a la población en general, la práctica habitual es aplicar un muestreo *por conglomerados polietápico*, que incluye la previa *estratificación* de la población de estudio:

1. La *población* se *estratifica por región y tamaño de hábitat*. En España se aplica la división político-administrativa en comunidades autónomas. La afijación del tamaño muestral suele hacerse de manera proporcional (al peso en el conjunto de la población).
2. Los *municipios* (o *puntos de muestreo*) se extraen al azar, con probabilidad proporcional al peso de la población de cada municipio. Para la selección de *puntos de muestreo* lo normal es utilizar datos agregados (frecuencias absolutas acumuladas de la población en cada estrato diferenciado por su tamaño poblacional) y la *tabla de números aleatorios*, como se ilustró en el ejemplo anterior y se muestra en el que sigue a continuación.

3. Selección *aleatoria* (en los municipios elegidos) *de áreas* o secciones estadísticas (como las *secciones censales*).
4. Elección de *viviendas* u hogares (habitualmente aplicando el muestreo por *rutas aleatorias*) e *individuos* (unidades últimas de muestreo) en las áreas seleccionadas. La elección de *individuos* se hará aplicando *muestreo por cuotas* u otro mencionado en el subapartado 7.3.6, ya sea de forma aislada o combinada.

En encuestas a la población en general, el nivel de agregación habitual suelen ser las *viviendas* u hogares. La unidad familiar o del hogar se presenta como un *conglomerado*, donde habrá que seleccionar (de manera aleatoria) uno o varios de sus integrantes. Se desaconseja la realización de más de una entrevista en un mismo hogar o vivienda, al preverse una mayor homogeneidad entre sus integrantes que entre los miembros de hogares diferentes. Recuérdese que, para la consecución de una muestra “representativa”, se prima la heterogeneidad como estrategia para reducir el *error muestral*.

EJEMPLO DE SELECCIÓN ALEATORIA DE PUNTOS DE MUESTREO

Ilustración de la selección de *puntos de muestreo* (municipios) en una encuesta a la población de 65 y más años de la Comunidad de Madrid, para conocer su calidad de vida. La población se restringe a la residente en hogares propios o ajenos (y no en residencias para la tercera edad u otros centros asistenciales). La encuesta se hará mediante entrevista *personal* en domicilios. Según el *Censo de Población y Vivienda de 2001*, en la Comunidad de Madrid residen 771.683 personas de 65 y más años. Para un *error muestral* máximo de $\pm 2\%$, un *nivel de confianza* del 95,5% y el supuesto de máxima *heterogeneidad poblacional* ($P = Q = 50$), al no disponer de información previa sobre la varianza poblacional en la variable de estudio, el tamaño de la *muestra* en el conjunto de la Comunidad de Madrid es de 1.600 unidades. Éstas se extraen de forma proporcional al peso de los municipios elegidos en el conjunto de la población de Madrid.

Los municipios de la Comunidad de Madrid se *estratifican* por tamaño de hábitat para proceder a la *afijación proporcional* de la *muestra*. Se segmenta la población en cinco estratos. Como el municipio de Madrid agrupa al 72,42% de la población de 65 y más años, podría plantearse una *estratificación no proporcional*, reduciendo el número de entrevistas a hacer en dicho municipio y así la muestra en los demás estratos. Pero, como en el caso más desfavorable, el estrato de menos de 1.000 habitantes, le corresponden 130 unidades muestrales por afijación proporcional, se opta por la *estratificación proporcional* (cuadro 7.15).

CUADRO 7.15. Determinación del número de puntos de muestreo según estratificación proporcional y distribución de la población de 65 y más años residente en la Comunidad de Madrid

<i>Tamaño del municipio por población de 65 y más años residente</i>	<i>Número de municipios</i>	<i>Población total de 65 y más años</i>	<i>Proporción que representa</i>	<i>Afijación proporcional</i>	<i>Número de puntos de muestreo</i>
Menos de 1.000 hab.	145	40.057	0,0519	130	8
De 1.000 a 5.000	21	47.501	0,0616	154	6
De 5.001 a 10.000	7	45.241	0,0586	147	3
De 10.001 a 50.000	5	80.022	0,1037	259	2
Más de 50.000	1	558.862	0,7242	1.810	1
Total	179	771.683	1,0000	2.500	20

El número de *puntos de muestreo* se determina considerando el número de municipios en cada estrato y su población. En el quinto *estrato* sólo hay un municipio, Madrid capital, lo que le exime de la selección aleatoria de municipios. Ésta sí se realiza en los cuatro restantes. Para ilustrar el proceso seguido en la *elección de los municipios* se toma el cuarto estrato (que agrupa municipios de 10.001 a 50.000 habitantes de 65 y más años), por ser el siguiente que reúne un menor número de municipios. En los demás estratos, el procedimiento sería el mismo. Primero, se listan los municipios incluidos en el estrato. Segundo, se calcula la *frecuencia absoluta acumulada* de su población (sumando la habida en los estratos precedentes). Después, se extraen tantos *números aleatorios* como *puntos de muestreo* a elegir (en el cuarto estrato sólo dos). De la tabla de *números aleatorios* (al final del capítulo), se extraen los dos primeros (de la primera columna) que cumplen la condición de ser igual o inferior a 80.022 (cantidad que representa a la población total agrupada en el estrato). En el cuadro 7.16 puede verse que, de acuerdo con las *frecuencias absolutas acumuladas*, los dos municipios elegidos (*unidades de muestreo* de segunda etapa) son los que primero incluyen al número aleatorio seleccionado: Móstoles y Alcorcón.

CUADRO 7.16. Elección aleatoria de los municipios (o puntos de muestreo) en el cuarto estrato

<i>De 10.001 a 50.000 habitantes</i>	<i>Población de 65 y más años</i>	<i>Frecuencia absoluta acumulada</i>	<i>Número aleatorio</i>
Móstoles	14.863	14.863	06318 47534
Getafe	15.617	30.480	
Alcalá de Henares	15.921	46.401	
Alcorcón	15.929	62.330	
Leganés	17.692	80.022	
Total	80.022		

El número de entrevistas en cada municipio será proporcional a su peso en el conjunto de la población. Partiendo de la hipótesis de que la calidad de vida puede estar afectada por la edad de la persona, la *muestra* se distribuye proporcionalmente en cada municipio según *cuotas* de género y edad. La distribución en los municipios de Madrid, Móstoles y Alcorcón se muestra en el cuadro 7.17.

CUADRO 7.17. Distribución proporcional de la muestra en los municipios del cuarto y quinto estratos por cuotas de género y edad

Municipios	Población total de 65 y más años	Varones			Mujeres		
		65-69	70-80	Más de 80	65-69	70-80	Más de 80
Móstoles	14.863	2.350	2.856	1.016	2.533	3.946	2.162
Alcorcón	15.929	2.890	2.932	1.074	2.994	3.785	2.254
Total	30.792	5.240	5.788	2.090	5.527	7.731	4.416
Madrid	558.862	73.878	107.118	39.048	94.640	157.775	86.403
Unidades muestrales							
Móstoles	125	20	24	9	21	33	18
Alcorcón	134	24	25	9	25	32	19
Total	259	44	49	18	46	65	37
Madrid	1.810	239	347	126	307	511	280

Como la encuesta se realiza en las viviendas, se seleccionan los hogares a partir de la elección, también aleatoria, de *secciones censales* (las unidades de muestreo de segunda etapa) por distrito. Alcorcón, *por ejemplo*, se halla seccionado en cuatro distritos con las siguientes secciones censales: 17, en el distrito 1; 34, en el distrito 2; 29, en el distrito 3; 28, en el distrito 4. Madrid cuenta con 21 distritos y las secciones censales van desde 27, en el distrito 21, hasta 217, en el distrito 10.

Después se concreta el número de entrevistas en cada sección. En encuestas a la población general, el CIS suele fijar un número alrededor de 10 entrevistas por sección (para propiciar una mayor heterogeneidad de la *muestra* conjunta). En el supuesto aquí ilustrado, al tratarse de una población más restringida (de 65 y más años), podría elevarse el número de entrevistas a hacer por sección. *Por ejemplo*, a 15, e inclusive 20, dependiendo de lo que se prime: ahorro de costes (en el *trabajo de campo*) o mayor precisión de las estimaciones de los parámetros poblacionales (que se alcanza a menor concentración de la *muestra* en unos *conglomerados* concretos).

La elección de los *individuos* (unidades de tercera etapa) se haría por *muestreo de áreas y rutas aleatorias* en cada sección elegida (subapartado 7.3.5). Tras la elección de la *vivienda*, procede la selección, también *aleatoria*, de la *persona* dentro de la vivienda, siguiendo alguno de los procedimientos incluidos en el subapartado 7.3.6.

El cuadro 7.18 resume las *ventajas e inconvenientes* principales del *muestreo aleatorio por conglomerados*.

CUADRO 7.18. Ventajas e inconvenientes del muestreo aleatorio por conglomerados

VENTAJAS	INCONVENIENTES
<ul style="list-style-type: none"> – Simplifica la elaboración del <i>marco muestral</i>: sólo exige un listado de los conglomerados elegidos – El <i>trabajo de campo</i> se concentra en un número limitado de <i>puntos de muestreo</i>, reduciendo los costes económicos y temporales de la investigación 	<ul style="list-style-type: none"> – Mayor <i>error muestral</i> (al haber más homogeneidad que heterogeneidad en los conglomerados*) – Requiere cálculos estadísticos complejos para la estimación del <i>error muestral</i> y de pesos para compensar diferentes probabilidades de selección y <i>errores de cobertura</i> – Para muestreos por conglomerados <i>polietápicos</i> precisa <i>marcos muestrales</i> específicos para cada etapa de selección de la muestra

* El *error muestral* se reduce aumentando el número de conglomerados. Los conglomerados grandes suelen ser más heterogéneos que los pequeños.

7.3.5. Muestreo de áreas y rutas aleatorias

Se aplican para la selección aleatoria de viviendas, cuando las unidades de muestreo son personas y familias; también de granjas, cosechas (en encuestas agrícolas), empresas u otras edificaciones. Su realización exige que haya *mapas* que cubran el *área* del ámbito espacial del estudio. Los *mapas* actúan como *marcos de muestreo*:

- La zona geográfica (país, comunidad autónoma, provincia, municipio) se divide en *áreas*, que constituyen *unidades de muestreo primarias*.
- Se eligen *áreas* de manera aleatoria.
- En las *áreas* elegidas, se crean nuevos *mapas* que ayuden a la elección aleatoria de *subáreas*, hasta concluir en la extracción (también aleatoria) de las *unidades* últimas de muestreo.

Para la *selección de viviendas*, los *mapas* han de especificar los límites de las *áreas* de la población de estudio. Se excluyen las *áreas* que no contienen viviendas: parques, colegios, iglesias, estaciones ferroviarias, de autobuses, hospitales, centros de internamiento... Aunque Kish (1965/1995) advierte que estos lugares pueden contener *viviendas* de las personas que están a su cuidado (trabajadores que pueden pertenecer a la población de estudio).

El *área* se divide en *bloques* porque contienen viviendas y tienen límites claramente identificados. Los bloques se estratifican según su similaridad para que en la *muestra* estén representados los diversos vecindarios del municipio. Se parte del supuesto de que los habitantes de un mismo vecindario comparten un perfil sociodemográfico similar. Además permite utilizar diferentes grados de conglomeración en distintos estratos. Los *bloques* se numeran y se les asignan medidas de tamaño, en función de las viviendas que incluyan. Si no se dispone de información, puede asignársele un número aproximado de acuerdo con el tamaño promedio de las viviendas en los vecindarios donde se procederá a la elección aleatoria de *viviendas*. A este respecto, el procedimiento más sencillo es la selección *sistemática* de los *bloques*, con el intervalo de elección aplicado a una lista de números de bloque. Este proceso de extracción muestral es, en palabras de Kish (1965: 308-309), “más fácil y adecuado” en la mayoría de los casos. Por el contrario, la selección “a pares” de bloques precisa más cuidado.

En las encuestas *personales* en domicilios, en las que se utilizan las *secciones censales* como unidades de segunda etapa (en un muestreo por *conglomerados polietápico*), la elección aleatoria de *viviendas* en las secciones elegidas se hace –como indica el que fuera responsable en el CIS, Rodríguez Osuna (1991)– de dos formas principales:

- a) Estudio detallado de las *secciones* (previo a la selección). El coordinador de la zona calcula (de forma aproximada) el número de *viviendas* y las refleja en un *croquis*. A partir de él, se fija la *ruta* y la frecuencia de selección de entrevistados: uno cada 100 viviendas, *por ejemplo*. Después, se extrae un *número aleatorio*, que es el que marca la *ruta* a seguir para la elección de la vivienda.
- b) El segundo procedimiento es el *método de rutas aleatorias*. Persigue la selección aleatoria de *viviendas* procurando mantener el principio de “equiprobabilidad”, presente en todo muestreo aleatorio.

Se denomina *muestreo de rutas* porque se establece la “ruta” o *itinerario* que el entrevistador ha de seguir para seleccionar las unidades de la muestra. Las *rutas* se eligen de forma “aleatoria” sobre un mapa del municipio donde se realizarán las entrevistas. En el *mapa* se indican múltiples puntos de partida o de inicio de *rutas*. El *punto de partida* suele corresponder a un edificio determinado por una dirección (*por ejemplo*, la calle Bravo Murillo n.º 13). En entidades de población pequeñas (núcleos rurales), el *punto de inicio* de la *ruta* suelen ser edificios centrales (el ayun-

tamiento, la iglesia, la plaza del pueblo...). La longitud de la *ruta* depende del número de unidades a extraer del *área* demarcada. Su origen es la dirección o punto de partida de la *ruta*. Su final, la última vivienda en la que se completa el número de entrevistas a hacer en dicha *área*. Tras la elección aleatoria del comienzo de la *ruta*, se especifican las *normas* para la elección de las unidades muestrales:

- Los *giros* a realizar (a la izquierda, a la derecha).
- Los *edificios* a entrar: los acabados en 2, *por ejemplo*.
- Si el *edificio* comprende más de una *vivienda*, la elección de la *escalera* (si hay más de una), de la *planta*, de la *vivienda* en la planta (siguiendo una tabla de números aleatorios específica).
- La *persona* a entrevistar en cada *vivienda* (de acuerdo con las *cuotas* fijadas u otro de los procedimientos indicados en el subapartado 7.3.6).

En las *hojas de ruta* el entrevistador registra sus actuaciones hasta localizar a las unidades de la *muestra* (y cualquier incidencia de utilidad para la supervisión del *trabajo de campo*). Suele incluir el número de entrevista, la dirección seguida (giro, calle o plaza, número), la selección de la *vivienda* (número total de plantas del edificio, planta elegida, número total de puertas en la planta, puerta elegida). Si la entrevista se llevó a efecto, la hora de realización, el sexo y la edad del entrevistado. En caso contrario, se anota el motivo de su no realización (*no respuesta por no contacto* o *por rechazo*).

EJEMPLO DE MUESTREO DE RUTAS ALEATORIAS

Criterios aplicados por el Centro de estudios de mercado socioeconómicos y de opinión, S. A. (Emopública), e incluidos en un *Manual del entrevistador* elaborado por dicho centro:

Dirección de partida

Corresponderá a un *edificio* determinado por una dirección (por ejemplo, c/ Capitán Haya n.º 65). En entidades pequeñas de población es un edificio concreto (*por ejemplo*, el ayuntamiento). En la localización de la dirección de partida pueden producirse las siguientes situaciones:

- a) Que el número designado como *punto de partida* no exista en la calle. En este caso, se tomará el edificio cuya numeración corresponda al primer número correlativo inferior que se encuentre (siendo éste par o impar, según el número dado como punto de partida).

- b) Que el número designado se encuentre repetido. Pueden darse tres casos:
- Números *bis*: se considera el número primitivo.
 - Números repetidos: el punto de partida será aquel número que ocupe el primer lugar, teniendo en cuenta el orden creciente de la numeración de la calle.
 - Numeración doble (antigua y moderna): siempre se considera la más moderna.
- c) Que termine la numeración de la calle antes de llegar al número dado como *punto de partida*. En este caso se tomará el edificio, dentro de la misma calle y acera, cuya numeración sea la más elevada.

Determinación de valores de “X”

Normalmente dos, que son previamente definidos por el departamento de campo. Se utilizan para seleccionar los *edificios* donde hacer las entrevistas a lo largo de la *ruta*. Por ejemplo, todo *edificio* cuya numeración sea 3, 7 o su terminación coincida con dichos números.

Seguimiento de la ruta

Para iniciar la *ruta* el entrevistador deberá colocarse de espaldas al *punto de partida* y comenzará a caminar hacia la derecha hasta encontrarse con la primera bocacalle, por la cual girará a la derecha (primer giro). Seguirá la *ruta* recorriendo la manzana correspondiente hasta encontrarse con la segunda bocacalle. En ella girará a la izquierda (segundo giro) y así sucesivamente (derecha-izquierda) hasta completar el número de entrevistas a realizar en la zona (figura 7.2).

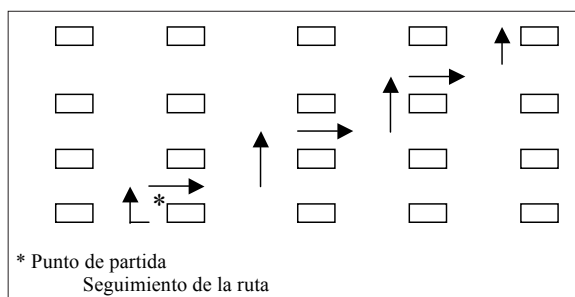


FIGURA 7.2. Croquis para el seguimiento de la ruta.

La *ruta* puede transcurrir por ambas aceras, según esté situada en ellas la numeración de las calles por las cuales va pasando. Algunas situaciones que pueden dar a confusión en la ejecución del itinerario son las siguientes:

- Que la calle por la que se avanza se bifurque. La *bifurcación* siempre se considera giro y éste se efectuará a la derecha o a la izquierda, según corresponda.

- Que en la *ruta* se encuentre una *plaza*. Al desembocar en ella, se procederá en función del orden de *giro* que corresponda efectuar. *Por ejemplo*, si es a la derecha, se correrá el tramo de la plaza hasta llegar a la primera bocacalle por la cual se saldrá de la plaza girando también a la derecha. Del mismo modo se procederá, pero a la izquierda, en caso de que haya que girar a la izquierda. Como norma general, en la *plaza* sólo se realizarán dos giros: el que corresponda a desembocar en ella y el que se realice para salir de la plaza.
- En caso de encontrarse, siguiendo la *ruta*, con un *descampado*, una valla prolongada o algún obstáculo que impida realizar el giro, el *giro* se hará en el sentido completamente opuesto al que se tenía que hacer. Después, se seguirá la *ruta* respetando el orden de *giros* (derecha e izquierda), a partir del último efectuado. A dicha operación se llama “inversión de ruta”.
- Si se trata de un *callejón* pueden darse dos situaciones:
 1. Que el callejón tenga salida a otra calle (figura 7.3). En este caso el callejón se recorrerá considerándolo como una bocacalle más en la *ruta*, y teniendo en cuenta el *giro* realizado a su entrada, exclusivamente;
 2. Callejón sin salida a otra calle (figura 7.4). Si es así, éste se recorrerá sin afectar en absoluto al orden de giros.

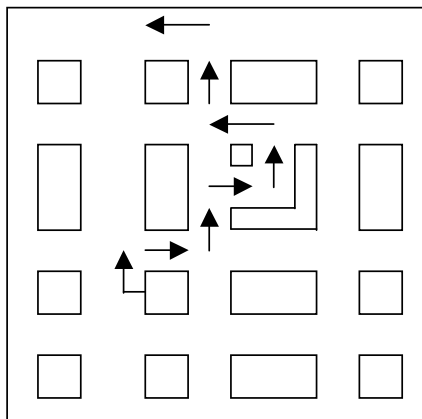


FIGURA 7.3. Callejón con salida a otra calle.

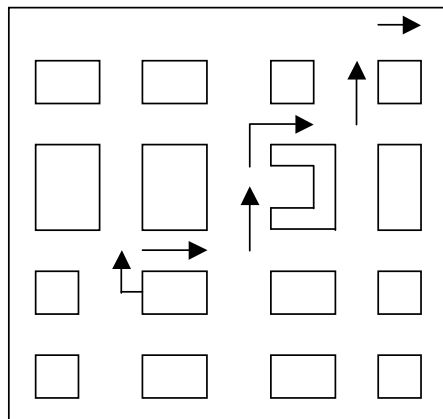


FIGURA 7.4. Callejón sin salida a otra calle.

Sistema de zona

Se emplea cuando las *viviendas* no forman calles con una disposición más o menos regular y no es posible seguir la *ruta* convencional.

- Si el *punto de partida* se encuentra en el interior del grupo de población o viviendas (figura 7.5). El entrevistador se situará en el edificio determinado como *punto de partida* (por ejemplo, el ayuntamiento). Comenzando en este punto se hará un recorrido envolvente (en espiral), alrededor de dicho punto de origen y en el sentido de la

aguja del reloj. De “dentro” a “fuera”, comenzando por la primera vivienda o edificio de la derecha.

- Si el *punto de partida* se encuentra en un lugar situado en el exterior del núcleo de población o viviendas (figura 7.6), se procederá del mismo modo, pero el recorrido de la espiral será ahora de “fuera” a “dentro”.

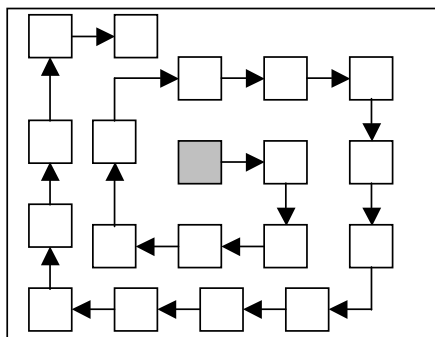


FIGURA 7.5. Punto de partida en el interior.

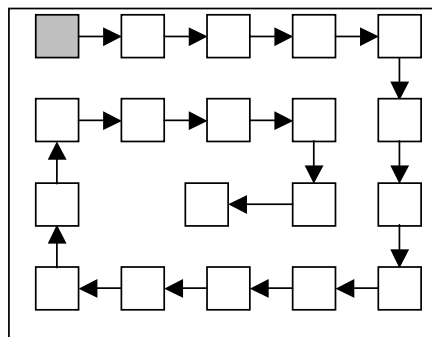


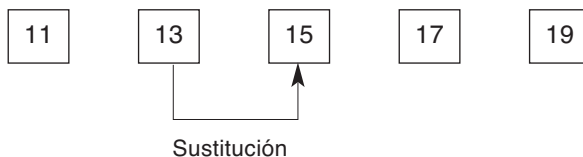
FIGURA 7.6. Punto de partida en el exterior.

La aplicación del “sistema de zona” se reflejará igualmente en la *hoja de ruta*.

Sustitución de edificios

Sólo cuando el entrevistador haya agotado todas las posibilidades para evitar su aplicación. Se trata de evitar alteraciones en la *muestra* que afecten a su representatividad. Como norma general, un edificio se sustituirá siempre por el siguiente edificio siguiendo la *ruta* en la misma acera, aunque la numeración del mismo no corresponda con los valores de X a aplicar.

Ejemplo, valores de X = 3 y 7.



La *sustitución del edificio* se hará cuando coincidan algunos de los casos siguientes:

- No sea un bloque de viviendas: instituciones (organismos oficiales, colegios, hospitales), hoteles, oficinas, comercios, fábricas y similares.
- Que sea una vivienda unifamiliar en la cual la totalidad de los miembros que la componen se encuentren ausentes, rehúsen colaborar o que se encuentre deshabitada.
- Un edificio de varias viviendas en el que, por diversos motivos (viviendas deshabitadas, hogares que rehúsan participar en la encuesta) no se haya podido hacer la entrevista en ninguna de las viviendas del edificio.

Selección de plantas en el edificio

Cuando tiene más de una planta, la selección de las *plantas* se hace utilizando una *tabla de números aleatorios* (o selector):

CUADRO 7.19. Tabla de números aleatorios para la selección de plantas en un edificio

		Número de la entrevista a realizar														
		1	2	3	4	5*	6	7	8	9	10	11	12	13	14	15
Número de plantas / puertas	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	2	1	2	1	2	1	2	2	1	1	2	1	2	2	1	2
	3	3	1	3	1	2	2	3	2	1	2	2	2	3	2	3
	4*	4	3	2	1	3*	3	2	2	1	4	1	4	4	1	3
	5	3	5	1	5	5	4	3	5	2	1	4	1	2	1	3
	6	6	4	3	5	2	1	3	4	2	1	3	5	4	3	6
	7	6	4	6	3	3	1	5	4	6	1	5	7	7	2	5
	8	2	4	8	2	6	7	2	5	5	1	3	8	7	6	2
	9	2	8	8	5	1	9	4	2	6	9	6	7	4	3	3
	10	5	8	8	5	9	7	5	1	3	6	4	7	2	6	9
		Planta o puerta seleccionada														

* Por ejemplo, Entrevista a realizar: quinta

5

Plantas en el edificio: cuatro

4

Planta seleccionada: tercera

3

Cruzando en la *tabla* el número de orden de la entrevista con el número de plantas del edificio o módulo seleccionado, se obtiene la *planta* elegida al azar dentro del edificio. Por “módulo” se entiende la agrupación de plantas dentro de un mismo edificio, cuando éste sobrepase cinco alturas.

- Es frecuente que en los edificios que tengan más de 5 *plantas* se realicen tantas entrevistas como módulos contenga el edificio. Para lo cual, se divide el número de plan-

tas habitadas que tiene el edificio en módulos de 5 plantas. Si el número de plantas no fuese múltiplo de 5, las últimas plantas restantes se considerarán, a efectos de selección, como un módulo más. Cada *módulo* se considera, a efectos de selección aleatoria de planta, un edificio “independiente”. En la *hoja de ruta* se anotará si el edificio contenía más de 5 plantas y qué plantas y puertas se seleccionaron.

- Si el edificio tuviese 2 o más escaleras, se procederá como si se tratase de un edificio con varios módulos. Se ordenarán las escaleras y se procederá a seleccionar las plantas correspondientes dentro de cada módulo.

Sustitución de plantas

Se llevará a cabo en la planta inmediatamente superior y así, sucesivamente, en el caso de que fueran necesarias nuevas sustituciones. Si se llegase a la última planta del edificio o módulo (o ésta hubiera correspondido inicialmente en la selección), la *sustitución* se hará en la primera planta del edificio o módulo. Si fuese necesario, se continuará el ciclo de sustitución hasta llegar de nuevo a la planta inicialmente seleccionada.

En caso de no haber podido realizar ninguna entrevista, se procederá a la *sustitución* por el siguiente edificio siguiendo la *ruta*. En él se seleccionará, de nuevo, planta y puerta.

Selección de viviendas

Si en la planta elegida hubiese más de una vivienda, habrá que proceder a la selección aleatoria de una *vivienda*. Para ello se utilizará el mismo selector que para la selección de *planta*: se cruzará el número de orden de la entrevista a realizar con el número de puertas (o viviendas) que tiene la planta seleccionada. Previamente, habrá que ordenar las puertas en las plantas.

- Si éstas estuviesen ya numeradas, se repetiría dicha numeración.
- Si la ordenación fuese alfabética, se traduciría en su equivalente numérico (*por ejemplo*, la letra A con el número 1; B, con el 2; y así sucesivamente).
- Si las puertas no tuviesen ningún tipo de numeración, habría que ordenarlas de izquierda a derecha, siempre en función del punto por el cual se ha llegado a la planta, teniendo como punto de referencia la escalera.

Sustitución de viviendas

Se hará cuando no sea una vivienda familiar (una oficina, pensión o similar). También, cuando se encuentre deshabitada o sus inquilinos rehúsen participar en la encuesta.

- Si la vivienda fuese *unifamiliar*, la sustitución se realizará en la siguiente vivienda (o edificio), siguiendo la *ruta*.
- En edificios o módulos de varias plantas, con *una única vivienda por planta*, la sustitución se hará en la vivienda de la planta inmediatamente superior.

- En edificios con varias plantas o módulos y *más de una vivienda por planta*, la sustitución se hará en la puerta inmediata de la derecha a la izquierda, según proceda en función de la ordenación que se haya dado a las mismas para su selección.

Toda *sustitución* que se realice deberá detallarse en la *hoja de ruta*. Para la *selección de las personas en la vivienda* se aplicará alguno de los procedimientos enunciados en el subapartado 7.3.6, y ajustándose las características de la persona a entrevistar (delimitadas en el diseño muestral).

La selección por *rutas aleatorias*, aunque garantice el principio de *equiprobabilidad* en la elección de las viviendas, no asegura que todas las unidades de la población tengan igual probabilidad de selección. Puede haber sobrerrepresentación (en la *muestra* final) de personas de mayor frecuencia en sus domicilios (amas de casa, jubilados, personas en paro, estudiantes). Se reducirá en la medida en que no se proceda a la rápida “sustitución” de los *no respondientes* (por no contacto o rechazo). Se recomienda que al menos se hagan tres *revisitas* (a la vivienda elegida al azar), en caso de *no contacto*, antes de su *sustitución*. Éstas y otras actuaciones se especifican en el apartado 7.5 (sobre el *error de no respuesta*).

7.3.6. Métodos de selección de individuos en una misma vivienda

Cuando haya que elegir uno o varios integrantes de una vivienda, existen varios procedimientos alternativos de selección que buscan la aleatoriedad: el método de “*no selección*”, de *selección aleatoria de Kish*, de *cuota sistemática* y del *cumpleaños*.

A) El método de “no”selección

El más deficiente de los métodos propuestos. No se considera un método *probabilístico*, aunque resulta poco intrusivo y de fácil realización. Se escoge a la primera persona que abra la puerta (encuesta *cara a cara*) o que conteste al teléfono (encuesta *telefónica*). La única condición es que se ajuste a la *población* de estudio (*por ejemplo*, persona de 18 años y más). Introduce “sesgos” en la muestra (mayor probabilidad de selección de personas que están más en casa, contestan al teléfono o abren la puerta). Para evitar su sobrerrepresentación, hay que emplear un método que equilibre la *muestra*, al menos en características sociodemográficas básicas (como género y edad), en igual proporción a su presencia en la *población* de estudio. Aplicar un método de selección *sistemático* por género (alternando varones y mujeres) podría corregir el *sesgo* por género, pero no por edad. Dada la mayor probabilidad a responder al teléfono, o a la puerta, de personas de menos edad, es preciso seguir un método “probabilístico” que ajuste la *muestra* al menos respecto a estas características básicas.

B) Método de Kish de selección aleatoria

Propuesto por Kish en 1949 y calificado por Lavrakas (1993: 108) como “el método de selección más riguroso que es aceptado para entrevistas en persona”. Su uso preferente es en encuestas *cara a cara* en domicilios, aunque también puede aplicarse en encuestas telefónicas. Si bien, en las *encuestas telefónicas* puede llegar a ser incluso más intrusivo que en la personal, con el consiguiente aumento de la *no respuesta* por rechazo.

Una vez que el entrevistador se ha presentado, solicita a la persona que responde (al teléfono, a la puerta) que enumere a los miembros del hogar que cumplan el criterio de edad de la población de estudio. *Por ejemplo*, personas de 18 a 55 años. Se excluye los que presenten alguna incapacidad (enfermedad mental, dificultad para la comunicación verbal...) y los ausentes por un largo período de tiempo. Se ordena a los que se ajustan a la *población* de estudio por su relación de parentesco con el cabeza de familia (cabeza de familia, esposa, padre del cabeza de familia, madre, hijo, hija, tía de la esposa u otro familiar en el hogar). Después, se pregunta por la edad de cada persona listada para ordenarlos de acuerdo con su género y edad. El varón de más edad será el número 1, el siguiente varón de más edad el número 2, y así de más a menos edad hasta incluir a todos los varones elegibles. Sigue la ordenación de las mujeres en la misma disposición, en orden de edad decreciente. Primero la mujer de más edad y, por último, la de menor edad. Después se aplica una *tabla* de números aleatorios, que se imprime junto al cuestionario (en su portada o anverso). Su composición se alterna al azar en los diferentes cuestionarios. Existen ocho versiones de la tabla de Kish. Su aplicación aleatoria se facilita bastante en los modos informatizados de encuesta (*CAPI*, *CATI*, *CASI* y *CAWI*).

Si la persona elegida no se halla en la vivienda, deberá preguntarse cuándo podrá localizarse e intentar concertar una cita. Lo que no debe hacerse es sustituirla por otro miembro del hogar que sí esté en ese momento y quiera ser entrevistado. Las negativas, de producirse (por no contacto o rechazo), deberán considerarse “entrevistas negativas” y proseguir con la selección aleatoria de viviendas.

EJEMPLO DE MÉTODO DE KISH DE SELECCIÓN ALEATORIA EN UNA MISMA VIVIENDA

Después de una breve presentación, el entrevistador solicita (a la persona con la que primero establece contacto) que liste a los miembros de la vivienda, en los términos siguientes:

“Para determinar a quién entrevistar, necesito listar a las personas de 16 y más años que habitualmente residen en esta vivienda. Sólo preciso conocer su edad y sexo, no sus nombres, y la relación de parentesco que tienen con usted. Comencemos por usted, ¿cuántos años tiene? ¿Y las otras personas?”

Con la información que proporciona se completa el cuadro 7.20.

CUADRO 7.20. Disposición de los miembros del hogar por relación con el cabeza de familia

<i>Relación con el cabeza de familia</i>	<i>Sexo</i>	<i>Edad</i>	<i>Adulto número</i>
Cabeza de familia	Varón	59	1
Esposo/a	Mujer	56	4
Madre del cabeza de familia	Mujer	82	3
Hijo	Varón	18	2
Hija	Mujer	21	5
Hija	Mujer	19	6
Sobrina	Mujer	17	7

Después, se aplica la *tabla* de selección propuesta por Kish (1949, 1965). La persona seleccionada es la que ocupa el número 4 (la esposa), de acuerdo con la siguiente tabla. Si no está en la vivienda, se concierta una cita (día y hora para la revisita).

CUADRO 7.21. Tabla de selección para elegir al encuestado

<i>Si el número de adultos en la vivienda es...</i>	<i>Entrevistar al adulto con el número</i>
1	1
2	2
3	2
4	3
5	4
6 o más	4*

* El adulto con el número 4 es el elegido al haber listadas 7 personas adultas.

El *método de Kish* propicia la “equiprobabilidad” en la selección de personas en una vivienda, pero presenta tres *inconvenientes* principales:

- Si existe sólo una persona elegible en el hogar, esa persona automáticamente se convierte en el sujeto seleccionado (Lavrakas, 1998). Obtiene *desiguales probabilidades de selección* para personas por tamaño del hogar: las que viven solas tienen una mayor probabilidad de ser elegidas frente a aquellas que viven con 6, 7 o más personas.
- Precisa que la persona que responda al teléfono o abra la puerta dé información de los otros miembros del hogar. Esta solicitud puede suscitar rece-

los y propiciar *no respuesta* por *rechazo* (dependerá de la destreza del entrevistador).

- c) El problema de la *veracidad de la información* que se proporcione de los otros miembros del hogar.
- d) Añade *minutos extra* al proceso de selección de entrevistados (más cuantas más personas habiten en la vivienda).

C) Método de cuota sistémica

Inicialmente propuesto por Trolldahl y Carter (1964), para ser menos invasivo y más rápido que el de Kish; y modificado por Bryant (1975). Dos son las preguntas a realizar:

- ¿Cuántas personas de X años o mayores viven en su casa incluyéndose Ud.?
- ¿Cuántos de ellos son varones?

Puede preguntarse por varones o por mujeres. La *edad* la marca la *población* de estudio. En la mayoría de los casos, la población de 18 años y más. La persona elegida resulta del cruce de la respuesta a ambas preguntas en una *tabla de selección* (cuadro 7.22). Se escoge una de las cuatro matrices o versiones de selección. Las versiones se aplican en el siguiente orden: 1.º, 2.º, 3.º, 4.º; 2.º, 3.º, 4.º; 1.º, 2.º, 3.º, 4.º; 2.º, 3.º, 4.º, etc. Se emplean de manera alternativa para que en cada contacto se aplique una diferente y consecutiva a la anterior, para garantizar una selección aleatoria.

Este método de selección (como el de *Kish*) puede alcanzar una muestra “equilibrada” (por género y edad) mezclando versiones de las *matrices de selección*. El inconveniente principal es el mismo: es un método “intrusivo”. Las dos preguntas que exige pueden considerarse personales. A ello se suma que muestra imprecisión en hogares con más de tres varones o tres mujeres.

D) Método del cumpleaños

Propuesto por Salmon y Nicholls (1983) y de amplia aplicación en encuestas telefónicas porque es un “método de selección superior a los demás”, según Lavrakas (1998: 448). Si bien, para Biemer y Lyberg (2003: 197) “no producen muestras aleatorias y no deberían utilizarse en la investigación científica rigurosa”.

Sólo precisa de una pregunta: qué persona (de los elegibles en el hogar) ha cumplido años “más recientemente” o lo hará “más próximamente”. *Por ejemplo*, “Para esta encuesta, ¿podría entrevistar a la persona de 18 años o más que vaya a cumplir años próximamente?”. Al reducirse las preguntas, su aplicación es más sencilla, rápida y menos intrusiva. Además propicia muestras *heterogéneas*, al darse a cada miembro de la vivienda igual probabilidad de selección.

CUADRO 7.22. Tabla para la selección aleatoria del entrevistado de Trol Dahl y Carter

Número total de varones en el hogar	Número total de personas de <u> X </u> años en el hogar			
	1	2	3	4 o más
<i>Versión 1</i>				
0	Mujer	Mujer mayor edad	Mujer menor edad	Mujer menor edad
1	Varón	Varón	Varón	Mujer mayor edad
2		Varón mayor edad	Varón menor edad	Varón menor edad
3			Varón menor edad	Varón mayor edad
4 o más				Varón mayor edad
<i>Versión 2</i>				
0	Mujer	Mujer menor edad	Mujer menor edad	Mujer mayor edad
1	Varón	Varón	Mujer mayor edad	Varón
2		Varón mayor edad	Mujer	Varón mayor edad
3			Varón menor edad	Mujer o mujer mayor edad
4 o más				Varón mayor edad
<i>Versión 3</i>				
0	Mujer	Mujer menor edad	Mujer mayor edad	Mujer mayor edad
1	Varón	Mujer	Varón	Mujer menor edad
2		Varón menor edad	Varón mayor edad	Varón mayor edad
3			Varón mayor edad	Varón menor edad
4 o más				Varón menor edad
<i>Versión 4</i>				
0	Mujer	Mujer mayor edad	Mujer mayor edad	Mujer menor edad
1	Varón	Mujer	Mujer menor edad	Varón
2		Varón menor edad	Mujer	Mujer menor edad
3			Varón mayor edad	Mujer o mujer mayor edad
4 o más				Varón menor edad

A estos métodos de selección se suman otros métodos, como el basado en “votantes probables” (de escasa aplicación, salvo en encuestas electorales): se pregunta, a la primera persona de la vivienda con la que se contacta (de 18 años o más), si votó en las elecciones pasadas; después, si piensa votar en las próximas elecciones. Si responde sí, se le entrevista; en caso contrario, no. También, puede preguntarse por una persona que se ajuste a unas *cuotas* concretas, de edad, género, etnia u otra característica que se considere relevante para el problema de investigación.

7.3.7. Diseños muestrales no probabilísticos

La selección de las unidades muestrales sigue criterios ajenos al azar: la tipicidad, la conveniencia o pertinencia, la oportunidad, disponibilidad o accesibilidad. Incluyen el muestreo *por cuotas*, el *estratégico* o de “juicio” y los *circunstanciales*. Este último incluye el llamado “bola de nieve”, el de *voluntarios* y el *ocasional*.

A) Muestreo por cuotas

De gran popularidad y aplicación en la investigación de mercados, favorecido por su menor coste (económico y temporal) y fácil administración. Fue inicialmente desarrollado Gallup y Ropper durante los años treinta, alcanzando un gran éxito tras su acierto en la predicción de los resultados de las elecciones presidenciales norteamericanas de 1936 (aunque erró en la correspondiente a 1948, como ya se vió). Para Kish (1965/1995: 563) es un método que carece de carácter de “científico”; para Lohr (2000: 115) se adecua “a la investigación comercial, pero no a otros tipos de encuesta”. Se debe a su condición de método *no probabilístico*. Si bien, dentro de los no probabilísticos es el método que proporciona mejores resultados. Al menos garantiza “proporcionalidad” de la *muestra* (respecto de la población de estudio) en las variables utilizadas como “cuotas”.

Al igual que el *estratificado*, el muestreo por *cuotas* parte de la segmentación de la población en grupos, en función de unas variables sociodemográficas concretas. Se buscan grupos internamente homogéneos y heterogéneos respecto de los otros grupos. Pero la selección de las unidades de la *muestra* no se realiza de manera *aleatoria* (a diferencia del *estratificado*), sino por libre elección. Sólo se exige que se ajuste a las “cuotas” marcadas. Para ello se elabora una *matriz* con las características básicas de la población a analizar (sexo y edad, nivel de instrucción, clase social...), a modo del muestreo *estratificado proporcional*. La información se obtiene de datos censales (censo de población, padrón de habitantes...). El objetivo es que la *muestra* se ajuste a la *población*. Tras la delimitación de las *cuotas*, se asigna a cada entrevistador la tabla con los perfiles a cubrir (de acuerdo con las *cuotas* marcadas).

EJEMPLO DE ASIGNACIÓN DE CUOTAS

En un estudio de actitudes hacia la inmigración se toma como variable clave la posición social. Se establecen *cuotas* por sexo, edad y clase social. A cada entrevistador se le asignan unas “cuotas” (a modo de la siguiente) para que contacte con personas que se ajusten a ellas:

CUADRO 7.23. Cuotas por sexos, edades y clases sociales

	CLASE SOCIAL			Total
	Alta	Media	Baja	
Varones				
18 - 29	1	2	3	6
30 - 44	1	2	4	7
45 - 64	2	4	5	11
65 y más	1	2	3	6
Mujeres				
18 - 29	1	2	5	8
30 - 44	1	2	6	9
45 - 64	2	4	4	10
65 y más	2	3	3	8
Total	11	21	33	65

El margen de libertad dado al entrevistador para la selección de la unidad de la muestra es la principal *debilidad* del muestreo por *cuotas*. Puede elegir a personas de su entorno, a aquellas más accesibles y de mayor propensión a colaborar en el estudio. Además es probable que, dentro de una misma *cuota*, se escoja a unos individuos con preferencia sobre otros. *Por ejemplo*, controlando por edad, elegir a más personas de edades intermedias en los intervalos diferenciados (y menos las situadas en los límites de los intervalos de edad). A lo que se suma el riesgo de que el entrevistador ubique a los sujetos en una *cuota* a la que realmente no pertenece (para completar casos sin cubrir). Una picaresca no fácil de controlar.

A estas debilidades se suma la general a los métodos *no probabilísticos*: el desconocimiento de las probabilidades de selección de las unidades de la población. Impide el cálculo de los *errores* muestrales. La *muestra* finalmente obtenida puede no representar a la *población* que se analiza, aunque se ajuste a ella en las *cuotas* fijadas. Éste es su principal inconveniente, al que se suman otros incluidos en el cuadro 7.24.

B) Muestreo estratégico o “de juicio”

Las unidades de la *muestra* se eligen de manera “estratégica”, siguiendo criterios o el “juicio” del investigador: creencia de que puede aportar información de “interés” (o relevancia) para los objetivos del estudio. Se precisa que sean informantes *heterogéneos* (que aporten puntos de vista diferentes sobre el problema de investigación) y *accesibles* (que quieran hablar).

CUADRO 7.24. Ventajas e inconvenientes del muestreo por cuotas

VENTAJAS	INCONVENIENTES
<ul style="list-style-type: none"> – Menor coste económico-temporal – Fácil proceso de selección de las unidades muestrales. No requiere un <i>marco</i> muestral nominativo 	<ul style="list-style-type: none"> – No puede precisarse el <i>error muestral</i> – Carece de base probabilística para la inferencia estadística – Propicia <i>sesgos</i> en la selección de las unidades de la <i>muestra</i> – Límites en la representatividad de la <i>muestra</i> para características no especificadas en las <i>cuotas</i> – Difícil control del <i>trabajo de campo</i>

Por ejemplo, escoger a alumnos que se piensa que representan diferentes corrientes de pensamiento y que pueden aportar información “valiosa” para los objetivos de la investigación. Su éxito dependerá del grado al que los informantes elegidos ex profeso sean realmente “representativos” de la heterogeneidad de la población de estudio. Antes de su aplicación, conviene especificar las dimensiones de variabilidad en la población que se consideran relevantes para los objetivos de la investigación.

Un muestreo adecuado para estudios de limitado presupuesto (como tesis doctorales u otras investigaciones académicas de escasos recursos económicos), enfocados a la comprobación de modelos teóricos, medición de conceptos... En *muestras* muy pequeñas (inferiores a 30 casos) el muestreo *estratégico* puede propiciar que se recabe información de mayor relevancia para los objetivos del estudio que la selección aleatoria. Pero, siempre que se escoja a informantes que representen diferentes vertientes del tema que se analice; y que el estudio no busque la representatividad “estadística”, sino de la “información” (captar la variedad opinática sobre el tema en cuestión). Como se verá en el apartado 7.4, en muestras aleatorias de muy pequeño tamaño la magnitud del *error muestral* se dispara, limitando las posibilidades de inferencia de los resultados de la investigación. Disminuye la probabilidad de que la *muestra* represente la heterogeneidad de la *población* de interés. No así conforme se incrementa el tamaño de la *muestra*.

C) Muestreos circunstanciales: de voluntarios y “bola de nieve”

Es la “accesibilidad” el criterio fundamental de elección de las unidades muestrales. Caso de investigaciones con estudiantes universitarios que se ofrecen “voluntarios” para participar en un experimento, o de sondeos realizados en programas de televisión o radio, o encuestas web cumplimentadas por personas que se ofrecen voluntariamente (y no son elegidas al azar), *por ejemplo*. Estudios de utilidad para

conocer diferentes opiniones sobre una determinada cuestión, pero no para medir su “representatividad” en el conjunto de la *población*. Los *voluntarios* no representan la diversidad de la población (y pueden ofrecer una visión distorsionada de la realidad).

Por ejemplo, una encuesta autocumplimentada por los huéspedes de un hotel no permite conocer el grado de satisfacción con la atención prestada porque son los huéspedes insatisfechos quienes más las complimentan (quieren dejar constancia de su “queja”). Su utilidad se restringe al conocimiento de motivos de queja para mejorar la atención a sus huéspedes. Salvo que todos los huéspedes cumplieran el cuestionario o una muestra aleatoria de ellos.

Esta categoría también incluye al muestreo *bola de nieve*: la *muestra* se va configurando a partir de las referencias de personas con las que ya se ha contactado. Poco a poco la muestra va ensanchándose como una “bola de nieve”, hasta que el investigador decide cortar. Generalmente, cuando se llega a la “saturación teórica”: la información comienza a ser redundante, no aportando ningún punto de vista analítico nuevo. Este tipo de muestreo es útil en estudios con poblaciones marginales o de difícil acceso.

Las ventajas e inconvenientes de estos muestreos *no probabilísticos* se resumen en el cuadro 7.25.

CUADRO 7.25. Ventajas e inconvenientes de los muestreos no probabilísticos

VENTAJAS	INCONVENIENTES
<ul style="list-style-type: none"> – Simplicidad y economía del diseño muestral – De fácil ejecución – No precisan un <i>marco</i> muestral 	<ul style="list-style-type: none"> – No cabe estimar el <i>error muestral</i> – Resultados de la investigación no generalizables al conjunto de la población, al ser la <i>muestra sesgada</i>

7.4. Determinantes del error muestral

Por muy perfecta que sea la *muestra*, como únicamente se analiza una parte de la *población* (y esa *muestra* sólo representa una de todas las posibles que pueden extraerse de la misma población), siempre habrá divergencias entre los valores obtenidos en la *muestra* (estimaciones muestrales) y los correspondientes a la *población* (parámetros poblacionales). Esa disparidad es el *error muestral* (o *error típico* de muestreo). Al igual que otros errores de *no observación* (capítulo 5), su magnitud depende del *tamaño* de la *muestra* y de su *representatividad*: si todos los componentes de la *población* han tenido igual probabilidad de ser elegidos para formar la *muestra*. Ha de evitarse su interpretación errónea como una medida de resumen de todos los errores de estimación de una encuesta. Su valor depende de:

- a) *El tamaño de la muestra*: a medida que aumenta, decrece el *error muestral* (cualquiera que sea el valor de la *varianza poblacional*). Aseveración que puede comprobarse en el cuadro 7.26, que reúne *errores típicos* correspondientes a diferentes tamaños muestrales. *Errores* que disminuyen conforme se incrementa la *muestra*, indistintamente de los valores de P y Q. Véase que reducir el *error* a la mitad supone cuadruplicar el tamaño de la *muestra*, por ejemplo.

CUADRO 7.26. Errores para diferentes tamaños muestrales a un nivel de confianza del 95,5%

Número de unidades en la muestra	Valores presupuestos de P y Q (%)					
	1/99	10/90	20/80	30/70	40/60	50/50
50	2,8	8,5	11,4	13,0	13,9	14,2
100	2,0	6,0	8,0	9,2	9,8	10,0
200	1,4	4,3	5,7	6,5	7,0	7,1
500	0,9	2,7	3,6	4,1	4,4	4,5
1.000	0,6	1,9	2,6	2,9	3,1	3,2
1.500	0,5	1,6	2,1	2,4	2,5	2,6
2.000	0,4	1,3	1,8	2,0	2,2	2,2
3.000	0,4	1,1	1,5	1,7	1,8	1,8
4.000	0,3	1,0	1,3	1,4	1,5	1,6
5.000	0,3	0,9	1,1	1,3	1,4	1,4
10.000	0,2	0,6	1,8	1,0	1,0	1,0
50.000	0,1	0,2	0,3	0,4	0,4	0,4

- b) La *varianza poblacional* en la variable de estudio: a mayor heterogeneidad, mayor *error* (como asimismo muestra el cuadro 7.26), por la menor probabilidad de que la *muestra* logre reunir toda la variedad que compone la *población*. Su reducción exige incrementar el tamaño de la *muestra*.
- c) El *nivel de confianza*: su aumento incrementa el tamaño de la *muestra* y reduce el *error muestral*. Aunque habrá que buscar el equilibrio entre precisión y coste. El *nivel de confianza* de mayor aplicación, 95,5%, expresa que hay un 95,5% de probabilidad de que la estimación del parámetro poblacional esté comprendida en el intervalo definido por dos veces el *error típico*.
- d) El *método de muestreo* aplicado: el *aleatorio estratificado* genera un menor *error muestral*, pero cuando existe heterogeneidad entre los estratos (y la estratificación responda a una variable clave en la investigación). El muestreo por *conglomerados* es (de los *probabilísticos*) el de mayor *error*. El *error* se incrementa a medida que aumenta el tamaño de los conglomerados en la última fase de muestreo (y la homogeneidad en ellos).

En suma, el *error muestral* se aminora conforme aumenta el *tamaño muestral* y la consiguiente representatividad de las unidades de la *población*. Precisa de la se-

lección aleatoria de las unidades muestrales (que se dé a todos los componentes de la *población* igual probabilidad de elección). El cuadro 7.27 ofrece las fórmulas de uso común para su estimación (corresponden al muestreo aleatorio *simple* y *sistemático*). Se diferencia entre *error típico de la media de Y* (variable cuya variabilidad se analiza) y de una *proporción (P)*; y entre *universo finito e infinito*. Para *poblaciones finitas* (de 100.000 unidades o menos), se aplica el *factor de corrección*:

$$\left(1 - \frac{n}{N}\right)$$

Si bien se considera aceptable ignorar dicho factor de corrección, cuando la *fracción de muestreo* ($f = n / N$) es menor de “0,1” (es decir, cuando menos del 10% de la población ha de ser muestreada). En *poblaciones infinitas*, la aplicación de este factor de corrección no es de utilidad, al ser la corrección prácticamente “1”. En ellas es el *tamaño de la muestra*, no el porcentaje de la población muestreada, lo que determina la precisión del estimador. Una *muestra* de 1.500 unidades tiene la misma precisión indistintamente de que la población esté formada por 100.000, 1.000.000 o más unidades.

CUADRO 7.27. Error típico para muestras aleatorias simples o sistemáticas

	POBLACIÓN INFINITA (> 100.000 unidades)	POBLACIÓN FINITA (≤ 100.000 unidades)
<i>Error típico de la media</i>	$Se_{(\bar{Y})} = \sqrt{\frac{\widehat{S}^2}{n}}$	$Se_{(\bar{Y})} = \sqrt{\frac{\widehat{S}^2}{n} \times \frac{N-n}{N-1}}$
<i>Error típico de una proporción</i>	$Se_{(\bar{P})} = \sqrt{\frac{\widehat{P} \times \widehat{Q}}{n}}$	$Se_{(\bar{P})} = \sqrt{\frac{\widehat{P} \times \widehat{Q}}{n} \times \frac{N-n}{N-1}}$

A partir del *error típico (Se)* se calculan los *intervalos de confianza*. Estos permiten graduar la precisión de las estimaciones muestrales. Expresan el intervalo de valores que comprende al parámetro poblacional a partir de la estimación muestral. Como toda inferencia estadística, los *intervalos* se estiman para un nivel de confianza concreto. Si es 99,7%, el *error* se multiplicará por 3; si es 95,5%, por 2; y si fuese 95%, por 1,96. El *intervalo* es igual al valor estimado en la *muestra* menos (límite inferior del intervalo) y más (límite superior del intervalo) el producto de *Z* por el *error típico de estimación (Se)*:

$$\text{Intervalos de confianza} = \bar{Y} \pm (Z \times Se_{(\bar{Y})}) \text{ o } \bar{P} \pm (Z \times Se_{(\bar{P})})$$

A mayor *error*, menor precisión de la estimación del parámetro poblacional, al ser más amplio el *intervalo de confianza*. Cuanto menor sea su amplitud, más se aproximará la estimación muestral al parámetro poblacional correspondiente. Los *intervalos de confianza* deberían informarse, junto con el estadístico muestral, para que se pueda apreciar la precisión de las estimaciones muestrales (o al menos precisar la magnitud del *error* muestral).

EJEMPLOS DE CÁLCULO DEL ERROR TÍPICO EN MUESTRAS ALEATORIAS SIMPLES Y SISTEMÁTICAS

- A) Una encuesta a 720 profesores de una universidad con un total de 6.450 docentes concluye que la edad media es de 47 años, con una desviación típica de 6,2 años. La *fracción de muestreo* es igual a 0,112 ($f = n / N = 720 / 6.450 = 0,112$); por lo que se aplica el *factor de corrección para población finita* ($N = 6.450$) y se calcula el *error de estimación* y los consiguientes *intervalos de confianza*:

$$Se_{(\bar{Y})} = \sqrt{\frac{\hat{S}^2}{n} \times \frac{N-n}{N-1}} = \sqrt{\frac{6,2^2}{720} \times \frac{6.450-720}{6.450-1}} = 0,218$$

Implica que la edad media de los docentes en dicha universidad (parámetro poblacional), con una probabilidad de acierto en la inferencia del 95,5%, estará comprendida entre 46,56 y 47,44 años. *Intervalo de confianza* que resulta de restar y sumar a la edad media muestral (47 años) el producto del *error* (0,218) por el valor Z que corresponde a un *nivel de confianza* del 95,5% ($Z = 2$): $\bar{Y} - (Z \times Se_{(\bar{Y})}) = 47 \pm (2 \times 0,218)$. Al ser mínima la amplitud del intervalo, implica que el *error* es muy pequeño, siendo muy elevada la precisión de la estimación muestral.

- B) Si la muestra de 720 docentes se extrae del conjunto de las universidades públicas y privadas en España, el *error* de estimación apenas varía, al igual que los *intervalos de confianza*, al ser baja la estimación de la varianza poblacional:

$$Se_{(\bar{Y})} = \sqrt{\frac{\hat{S}^2}{n}} = \sqrt{\frac{6,2^2}{720}} = 0,231$$

La edad media estaría comprendida entre 46,54 y 47,46 años, a un *nivel de confianza* del 95,5% ($Z = 2$): $\bar{Y} \pm (Z \times Se_{(\bar{Y})}) = 47 \pm (2 \times 0,231)$

- C) El *error* sería, en ambos casos, superior respecto de la estimación muestral de un 56% de satisfechos con la calidad de la enseñanza impartida: $\pm 1,744$ y $1,850\%$, respectivamente. Al igual que los *intervalos de confianza*, que van de 52,51 al 59,49% y del 52,3% al 59,7%, respectivamente, $[\hat{P} \pm (Z \times Se_{(\hat{P})}) = 56 \pm (2 \times 1,744) ; \hat{P} \pm (Z \times Se_{(\hat{P})}) = 56 \pm (2 \times 1,850)]$,

para una probabilidad de acierto en la inferencia del 95,5%. Ampliación que responde a la menor precisión de la estimación muestral, al ser elevada la *varianza poblacional* estimada.

$$Se_{(\hat{P})} = \sqrt{\frac{\hat{P} \times \hat{Q}}{n} \times \frac{N-n}{N-1}} = \sqrt{\frac{56 \times 44}{720} \times \frac{6.450-720}{6.450-1}} = 1,744$$

$$Se_{(\hat{P})} = \sqrt{\frac{\hat{P} \times \hat{Q}}{n}} = \sqrt{\frac{56 \times 44}{720}} = 1,850$$

En muestras *aleatorias estratificadas de manera proporcional*, se introducen las siguientes modificaciones en el cálculo del *error muestral*:

- *Error típico de estimación de la media*: $Se_{(\bar{Y}_{prop})} = \sqrt{\frac{\sum_{i=1}^n n_i \times \hat{S}_i^2}{n^2}}$
- *Error típico de estimación de una proporción*: $Se_{(\hat{P}_{prop})} = \sqrt{\frac{\sum_{i=1}^n n_i \times \hat{P}_i \times \hat{Q}_i}{n^2}}$

Donde: “ \hat{P}_i ” es la proporción de la *muestra* en el *estrato* i que posee el atributo en cuestión.

“ $\hat{Q}_i = 1 - \hat{P}_i$ ”, la proporción que no posee dicho atributo.

“ $\hat{S}_i^2 = \hat{P}_i \times \hat{Q}_i$ ”, estimación de la *varianza* de la variable de interés para la población en el *estrato* i.

“ $\sum_{i=1}^n$ ”, sumatorio de todos los *estratos*, desde 1 hasta n.

“ n_i ”, tamaño de la *muestra* en el *estrato* i.

“ n ”, tamaño de la *muestra* total.

EJEMPLO DE CÁLCULO DEL ERROR TÍPICO EN MUESTRA ALEATORIA ESTRATIFICADA

Una encuesta a los alumnos de las Facultades de Ciencias Sociales y Jurídicas de la Universidad Complutense concluye que el 69% de los 980 entrevistados es contrario a la reforma universitaria. Los porcentajes varían por centros. La *muestra* se afijó de manera proporcional al número de alumnos matriculados en cada facultad. Para ello se multiplica la proporción que representa cada uno de los *estratos* (facultades), columna (2) en el cua-

dro 7.28, por el *tamaño muestral* (980), obteniéndose las cifras en la columna (3) del cuadro siguiente. Para el cálculo del *error muestral* se parte de los porcentajes de alumnos contrarios a la reforma obtenidos en cada facultad (columna 4).

CUADRO 7.28. Distribución proporcional de la muestra por estratos (facultades universitarias)

Facultades	Población (1)	Proporción población (2)	n_i (3)	P_i (4)	$n_i P_i (1 - P_i)$ (5)
Ciencias de la Información	6.954	0,251	246	0,77	43,57
Derecho	9.114	0,329	322	0,55	79,70
Documentación	483	0,017	17	0,58	4,14
Económicas y Empresariales	4.518	0,163	160	0,68	34,82
Políticas y Sociología	3.100	0,112	110	0,81	16,93
Psicología	3.515	0,127	125	0,75	23,44
Total	27.684	0,999	980	0,69	203,60

El *error* se calcula multiplicando el número de alumnos en la *muestra* en cada estrato (columna 3) por la proporción de alumnos contrarios a la reforma universitaria (columna 4) y los partidarios ($\hat{Q}_i = 1 - \hat{P}_i$), en cada estrato. El resultado figura en la columna (5). Se suman dichos productos y se sustituyen en la fórmula siguiente:

$$Se_{(P_{prop})} = \sqrt{\frac{\sum_{i=1}^n n_i \times \hat{P}_i \times \hat{Q}_i}{n^2}} = \sqrt{\frac{203,60}{980^2}} = 0,0146$$

En este caso el *error* es muy bajo y similar al obtenido si la muestra hubiese sido aleatoria *simple*:

$$Se_{(\hat{P})} = \sqrt{\frac{\hat{P} \times \hat{Q}}{n} \times \frac{N - n}{N - 1}} = \sqrt{\frac{0,69 \times 0,31}{980} \times \frac{27.684 - 980}{27.684 - 1}} = 0,0150$$

Elevando ambos *errores* al cuadrado (*varianza*), se comprueba el *efecto del diseño* (*deff*), propuesto por Kish (1965/1995):

$$deff = \frac{(0,0146)^2}{(0,0150)^2} = 0,947$$

Implica que la *varianza* de la muestra *estratificada proporcionalmente* es un 5,3% más pequeña que la *varianza* de la muestra *aleatoria simple*. Por consiguiente, para alcanzarse

igual precisión en la estimación de los parámetros poblacionales mediante muestreo aleatorio *simple*, habría que ampliar el *tamaño muestral* a 1.035 (esta cifra se obtiene dividiendo 980/0,947). De esta forma se demuestra la mayor precisión alcanzada aplicando *muestreo aleatorio estratificado*. La diferencia entre ambos *errores* aumenta conforme se incrementan las variaciones observadas en los estratos respecto a la variable considerada.

En el muestreo aleatorio *por conglomerados*, el *error típico* depende de la correlación de los valores entre los integrantes de un mismo conglomerado y del número de unidades elegidas en el conglomerado. La *correlación intraconglomerado* mide la covariación entre los pares de unidades en el mismo conglomerado. Se calcula por desviaciones de la media global. La fórmula correspondiente al *error típico* para la estimación de la *media* poblacional sería:

$$Se_{(\bar{Y}_{congl})} = \sqrt{\left(1 - \frac{m}{M}\right) \times \frac{Sb^2}{m}}$$

Donde: “*M*” es el número de *conglomerados* en la *población*.

“*m*”: el número de *conglomerados* seleccionados en la *muestra*.

“*Sb*²”: la *varianza* de los valores del conglomerado *x_i*. Se obtiene:

$$Sb^2 = \frac{1}{m-1} \sum_{i=1}^n (\bar{Y}_i - \bar{Y})^2$$

Las estimaciones de los *errores típicos* para muestras aleatorias por *conglomerados* (sobre todo *polietápicas*) resultan bastante complejas. No obstante, existen varios programas estadísticos (SAS, Stata, Wesvar o Sudaan) que calculan los *errores típicos* para cualquier variedad de diseño muestral. Por último, insistir que el *error muestral* es sólo uno de los integrantes del *error total de encuesta*, aunque sea el que siempre se referencia, a diferencia de los otros errores descritos en el cap. 5.

EJEMPLO DE FICHA TÉCNICA

Sondeo Preelectoral Elecciones Generales, 2011 – CIS (Estudio n.º 2.915 - octubre 2011):

Ámbito: Nacional.

Universo: Población española de ambos sexos de 18 años y más.

Tamaño de la muestra: Diseñada: 17.397 entrevistas / Realizada: 17.236 entrevistas

La distribución provincial se incluye en el cuadro adjunto al final de la ficha técnica.

Afijación: No proporcional.

Ponderación: Para tratar la muestra en su conjunto es necesaria la aplicación de los coeficientes de ponderación que figuran en el cuadro al final de la ficha técnica.

Puntos de muestreo: 1.357 municipios en 50 provincias y las Ciudades Autónomas de Ceuta y Melilla.

Procedimiento de muestreo: polietápico, estratificado por conglomerados, con selección de las unidades primarias de muestreo (municipios) y de las unidades secundarias (secciones) de forma aleatoria proporcional, y de las unidades últimas (individuos) por rutas aleatorias y cuotas de sexo y edad.

Los *estratos* se han formado cruzando las 17 comunidades autónomas con el tamaño de hábitat, dividido en 7 categorías: menor o igual a 2.000 habitantes; de 2001 a 10.000; de 10.001 a 50.000; de 50.001 a 100.000; de 100.001 a 400.000; de 400.001 a 1.000.000, y más de 1.000.000 de habitantes.

Los *cuestionarios* se han aplicado mediante entrevista personal en los domicilios.

Error muestral: Para un nivel de confianza del 95,5% (dos sigmas), y $P = Q = 0,50$, el error es de $\pm 0,76\%$ para el conjunto de la muestra y en el supuesto de muestreo aleatorio simple.

Los *errores* que corresponden a cada provincia se encuentran en el cuadro al final de la ficha técnica.

Fecha de realización: Del 6 al 23 de octubre de 2011.

CUADRO 7.29. Extracto del cuadro de puntos de muestreo, entrevistas y coeficientes de ponderación*

Provincias	Municipios	Entrevistas		Error (%)	Ponderación
		Diseñada	Realizada		
Álava	13	378	374	5,17	0,331
Albacete	20	270	269	6,10	0,571
Alicante/Alacant	42	426	424	4,86	1,428
Almería	21	270	267	6,12	0,827
Ávila	24	255	251	6,31	0,271
Badajoz	27	270	270	6,09	1,030
Balears (Illes)	30	357	356	5,30	1,016
Barcelona	56	810	797	3,54	2,453
Burgos	16	270	269	6,10	0,545
Cáceres	24	270	270	6,09	0,636
Cádiz	31	357	357	5,29	1,365
.....
Valencia	38	434	425	4,85	2,200
Valladolid	9	270	268	6,11	0,799
Vizcaya	32	417	417	4,90	1,120
Zamora	21	255	249	6,34	0,324
Zaragoza	16	355	355	5,31	1,021
Ceuta	1	200	200	7,07	0,148
Melilla	1	200	193	7,20	0,129
Total	1.359	17.397	17.236	0,76	

*En el Banco de Datos del CIS puede consultarse la ficha completa

7.5. El error de “no respuesta” y su reducción

El diseño de una *muestra* “perfecta” no garantiza la plena participación de todas sus unidades. La *muestra* final puede diferir de la diseñada, sesgando los resultados de la investigación. Los datos extraídos de la *muestra* pueden alcanzar significatividad estadística (porque se mantiene el tamaño de la *muestra*), y a su vez no “representar” a la población de estudio (si no se mantiene la composición de la *muestra* original e intervienen otros factores ajenos a la *aleatoriedad* en su selección). La *no respuesta* limita la generalización de los hallazgos de la investigación. Su efecto depende de la magnitud de la *no respuesta* y de su composición: las características que diferencien a quienes sí responden de aquellos que no aportan información (más cuando dichas diferencias están directamente relacionadas con el problema de investigación). Por lo que se precisa conocer la composición de la *muestra* final y su correspondencia con la *población* de estudio.

La *no respuesta* puede ser *total* (o de unidad) y de *ítem* (o pregunta). La primera puede deberse a “no contacto” (con la unidad de la *muestra*) o a “rechazo” (ésta rehúsa participar en el estudio por deseo propio, incapacidad o problemas para la comunicación verbal). La segunda, la *no respuesta de ítem* (no se responde a preguntas del cuestionario) puede deberse al *cuestionario* (es demasiado largo, incluye muchas preguntas abiertas, preguntas *filtro* y *rut*as no bien delimitadas...), al *encuestado* (falta de conocimiento, decisión, inhibición...) o al *entrevistador* (no formula la pregunta o anota la respuesta). A diferencia de la *no respuesta de unidad*, la de *ítem* se detalla en una encuesta (“no sabe”/ “no contesta”), para cada pregunta, y se considera para la interpretación de los datos de encuesta. Para su estimación se analizan las unidades de la *muestra* que no han aportado información: los “no sabe” (salvo que sea ésta una categoría de respuesta relevante a la pregunta), “no contesta”, y los “no aplicables” (personas a las que no correspondía hacer la pregunta al haber una pregunta *filtro*). Se suman y dividen entre el total de la muestra analizada.

$$\text{Tasa de no respuesta de ítem} = \frac{\text{Encuestados que no aportan información}}{\text{Muestra analizada}} = \frac{\text{NS} + \text{NC} + \text{NA}}{\text{NS} + \text{NC} + \text{NA} + \text{R}}$$

Donde: NS = no saben.

NC = no contestan.

NA = no aplicables.

R = responden a la pregunta de forma “válida”.

En la *no respuesta total* se considera al conjunto de las unidades de la *muestra* original que no han acabado participando en el estudio:

$$\text{Tasa de no respuesta total} = \frac{\text{No participantes en la encuesta}}{\text{Muestra original}} = \frac{\text{NE} + \text{NC} + \text{N}}{\text{NE} + \text{NC} + \text{N} + \text{R}}$$

Donde: NE = unidades muestrales no elegibles o entrevistables, al no ajustarse a la población de interés.
 NC = unidades muestrales no contactadas.
 N = unidades muestrales que se han negado a participar en el estudio.
 R = unidades muestrales que participan o responden.

El análisis de la *no respuesta* no queda circunscrito a las *tasas de no respuesta*. También incluye la comparación entre quienes aportan información y los no respondientes. Esto último de más difícil conocimiento, salvo que se realice una investigación ex profeso. El *error de no respuesta* queda así definido:

Error de no respuesta = (tasa de no respuesta) \times (Diferencia entre la media de Y en los respondientes y no respondientes: $\bar{Y}_R - \bar{Y}_{NR}$).

El *error de no respuesta relativo* (o porcentual) permite comparar variables en diferentes niveles de medición (como en el error de *cobertura*):

$$\text{Error de no respuesta relativo} = t_{NR} \times \left(\frac{\bar{Y}_R - \bar{Y}_{NR}}{\bar{Y}} \right)$$

Donde: t_{NR} = tasa de *no respuesta total*.

\bar{Y}_R = media de Y (la variable analizada) en los respondientes.

\bar{Y}_{NR} = media de Y en los no respondientes.

\bar{Y} = media de Y en toda la muestra: $\bar{Y} = (t_R \times \bar{Y}_R) + (t_{NR} \times \bar{Y}_{NR})$

EJEMPLO DE CÁLCULO DEL ERROR DE NO RESPUESTA

Una encuesta *por correo electrónico* (EMS) a 1.500 estudiantes universitarios concluye que la calificación media del expediente académico es 7,8 (\bar{Y}_R). Si bien sólo 855 estudiantes aportaron información. La tasa de *no respuesta* fue del 43%. Una encuesta telefónica posterior a los *no respondientes* (645) obtiene una calificación media de 6,4 (\bar{Y}_{NR}). El error de *no respuesta* sería igual a 78,59 euros:

$$E_{NR} = t_{NR} \times (\bar{Y}_R - \bar{Y}_{NR}) = 0,43 \times (7,8 - 6,4) = 0,602$$

Para su comparabilidad, se calcula el *error de no respuesta relativo*:

$$\bar{Y} = (t_R \times \bar{Y}_R) + (t_{NR} \times \bar{Y}_{NR}) = (0,57 \times 7,8) + (0,43 \times 6,4) = 7,198$$

$$\text{Error no respuesta relativo} = t_{NR} \times \left(\frac{\bar{Y}_R - \bar{Y}_{NR}}{\bar{Y}} \right) = 0,43 \times \left(\frac{7,8 - 6,4}{7,198} \right) = 0,084$$

Para una *tasa de no respuesta* del 43%, el *error de no respuesta relativo* es de 8,4%.

La información de los *no respondientes* puede extraerse de:

- Los propios *no respondientes*, su seguimiento en un estudio ex profeso.
- *Imputándoles características de los respondientes* más difíciles de contactar (han precisado un mayor seguimiento: rellamadas, visitas, reenvío de cuestionarios...).
- *Preguntando a personas “próximas”* al no respondiente (vecinos, familiares, amigos, compañeros), que puedan aportar información sobre él (sexo, edad, estudios, ocupación, estado civil, número de personas en la vivienda...).
- *Extrayendo información del marco muestral* que se ha utilizado, aunque esta información suele ser escasa y no relacionada con los objetivos del estudio.
- *Comparando las características sociodemográficas de los respondientes con los datos censales* disponibles de la población de estudio. “Éste es por ahora el método más frecuente para asegurar los sesgos de no respuesta” (Groves, 1989: 187). Si bien, está el problema de su actualización (que no sean datos “viejos”) y de su “fiabilidad” (su correspondencia con la realidad).

Para la reducción del *error de no respuesta* pueden seguirse alguna o varias de las actuaciones siguientes:

- a) *Adecuar el trabajo de campo a la población* a analizar: evitar fechas u horas “atípicas”, en las que se prevean “ausencias” en los hogares u espacios donde se realice el estudio. El *tiempo del trabajo de campo* puede ampliarse para posibilitar el seguimiento de los “no contactados”, antes de su *sustitución*.
- b) *Preaviso de la realización del estudio* para reducir la *no respuesta* por “rechazo”. El preaviso puede hacerse por *correo* o *teléfono*. El *teléfono* permite concertar una cita (día y hora); responder a cualquier duda; sondear la predisposición a participar; y comprobar la adecuación del *marco muestral* (que se contacte con la persona registrada). Inconvenientes: tener los números de teléfono de todas las unidades de la *muestra*; destinar una parte del presupuesto y de personal a su realización (eleva los costes en tiempo y dinero); y puede poner en sobreaviso a la persona dándole tiempo para meditar su participación en el estudio.

La *prenotificación por carta* es más habitual y se asocia a un mayor aumento de la tasa de respuesta. Se recomienda que la carta incluya el *membrete* del patrocinador del estudio, que esté *personalizada* (dirigida a una persona concreta, con su nombre y apellidos, y firmada por el responsable del estudio) y redactada en un *lenguaje sencillo* que motive a la participación (destacando la importancia y beneficios de la colaboración). A la carta puede añadirse un *incentivo* económico-material *ipso facto* o prometido.

- c) *Seguimiento de la muestra* mediante rellamadas, visitas o reenvío del cuestionario (en encuestas por *correo*). De mayor aplicación, por su menor coste, en la encuesta *telefónica*, y más restrictiva en estudios mediante entrevista *personal*

por su mayor coste añadido. Kish (1965/1995) recomienda repetir la visita hasta seis veces o más, hasta que se alcance una *tasa de respuesta* del 85%.

- d) *Gratificar la participación*: es preferible un regalo “simbólico”, de escaso valor (cinco euros adjunto al cuestionario, *por ejemplo*) que una *gratificación* cuantiosa (un coche, 6.000 euros, un viaje para dos personas...) “prometida” tras su cumplimentación (Berk *et al.*, 1987; Mangione, 1998; Singer, Van Hoewyk y Maher, 2000). Las *gratificaciones* se utilizan mucho en *encuestas panel* y de *diario* para contribuir a que la persona continúe participando en el estudio. En su contra, el incremento en los costes de la investigación.
- e) *Mejorar la presentación del estudio*: resaltar los beneficios de la participación y garantizar el *anonimato* y la *confidencialidad* de la información dada.
- f) *Emplear otro método de encuesta* en los no respondientes: la encuesta *cara a cara* a quienes no han devuelto el cuestionario por *correo* o no han podido contactarse con ellos por *teléfono* (cuando se dispone de su dirección), *por ejemplo*. Además, su uso complementario ayuda a recabar información del perfil de los no respondientes. Información que se precisa para calibrar el *error de no respuesta* y consiguiente “representatividad” de la *muestra* analizada.
- g) *Convertir las “negativas” en aceptación*: reasignando los casos “reacios” a entrevistadores con mayor experiencia y formación, enviando una *carta* de “persuasión” que reitere la importancia de la participación, o cambiando el método de encuesta, *por ejemplo*.
- h) *Incrementar el tamaño de la muestra*, siempre y cuando no se traduzca en captar a las personas más fáciles de “contactar”.
- i) *Sustitución por unidades equiparables*. Junto con la *imputación*, es una de las actuaciones más frecuentes para reducir la *no respuesta*. Si bien, puede generar un mayor *error*, si los “sustitutos” no se equiparan a quienes reemplazan. Por su efecto en la pérdida de representatividad de la *muestra*, debería ser el último recurso a aplicar. Es de mayor aplicación en la investigación comercial.
- j) *Ajustes estadísticos mediante la ponderación e imputación* para equilibrar las diferencias en las estimaciones de los respondientes y no respondientes. Precisa de información sobre los *no respondientes*; al menos, sus características sociodemográficas básicas (género, edad, estatus social, hábitat...). Aunque no siempre se dispone de esta información.

El procedimiento de corrección de la *no respuesta* mediante la *ponderación* es similar a su aplicación en la *estratificación no proporcional* (subapartado 7.3.3). Se incrementa el *peso* de los respondientes para que “representen” a los que no han respondido mediante alguno de los procedimientos habituales siguientes:

- *Ponderar de acuerdo con la distribución de la población*. No requiere información sobre los *no respondientes*. Se “ajustan” características demográficas de la *muestra* final respecto a las mismas características en la población de interés, a partir de datos censales disponibles de la población.

Datos sociodemográficos (género, edad, estudios, ocupación, hábitat...) no siempre están relacionados con el problema de investigación. En cada estrato, se divide el porcentaje que representa en la *población* entre el porcentaje que representa en la *muestra*. Al igual que en el muestreo *estratificado no proporcional*, cada peso o *coeficiente de ponderación* se multiplica por la estimación muestral correspondiente en cada grupo de *población*. Después, se suman todos los productos y se divide entre el total de la *muestra* analizada. Resulta la estimación para el conjunto de la *población*, corregida la desproporcionalidad en la *muestra* (por la incidencia desigual de la *no respuesta*).

Esta *ponderación* quiere corregir errores de *no respuesta* y de *cobertura*. Su principal *inconveniente* es que sólo puede hacerse respecto a variables (en su mayoría sociodemográficas) de las que se tenga información, con independencia de su relevancia respecto al problema de investigación. Su éxito dependerá de su correspondencia con los objetivos del estudio.

- *Ponderar según la muestra original*: se “ajusta” la *muestra* final a la inicial. El *coeficiente de ponderación* se obtiene dividiendo la *muestra* original entre la *muestra* final, en cada estrato o grupo diferenciado (de acuerdo a una o más variables de clasificación). Su precisa conocer a qué grupo de población pertenecen tanto los *respondientes* como los *no respondientes*. A menos que esta información se recabe ex profeso, la práctica usual es utilizar la empleada para el diseño de la muestra, en su estratificación (género, edad, estudios, ocupación, hábitat). La *ponderación* será más exitosa, cuanto más relacionadas estén las variables elegidas con el problema de investigación.
- *Ponderar según la accesibilidad de las unidades de población: método de Politz-Simmons*. A la *muestra* se pregunta cuántos días, de los últimos seis, se encontraba en su domicilio a la misma hora. La *probabilidad* de estar en casa se obtiene dividiendo: $(K + 1) / 6$. Siendo “K” el número de tardes de las cinco precedentes. Una persona que haya estado en casa todos los días tendrá probabilidad máxima $[(5 + 1) / 6 = 1]$; quienes no hayan estado ningún día, probabilidad mínima $[(0 + 1) / 6 = 0,1667]$. La *muestra* se estratifica en seis grupos definidos por la “probabilidad de estar en casa” (en la franja horaria para el *trabajo de campo*): $6/6 = 1$; $5/6 = 0,833$; $4/6 = 0,667$; $3/6 = 0,5$; $2/6 = 0,333$; $1/6 = 0,1667$. Después se pondera por su inversa (se parte del supuesto de que los menos accesible se asemejan a los finalmente “no contactados”).

Por ejemplo, $1 / 0,1667 = 5,999$. O, igualmente por $6/6$, $6/5$, $6/4$, $6/3$, $6/2$ y $6/1$. Son los pesos o *coeficientes de ponderación*. Quienes dicen haber estado en casa todos los días tendrán un coeficiente menor (1) y el máximo (5,999) aquellos que no han estado ningún día. Los coefi-

cientes se multiplicarán por las estimaciones muestrales, al igual que las ponderaciones precedentes.

Pero, salvo que se haga un estudio *ex profeso*, no hay plena certeza de que los *no respondientes* se asemejen a los *respondientes*. Se suma, además, el problema de la *fiabilidad* de la información que precisa: el número de días que estaba en casa (la voluntad de informar y la exactitud del recuerdo).

k) Respecto a la *imputación* (“asignación de valores” a los casos en los que faltan), los procedimientos de uso más común son:

- *Imputar por la media*: “reemplazar” los valores *sin respuesta* por la *media* de los valores observados antes de proceder al análisis.
- *Imputar por otros casos “similares”*: a los casos *sin respuesta* se les atribuye la dada por otros individuos de similares características, que han dado iguales respuestas que ellos en otras preguntas. Se la llama “*hot deck*” porque duplica la respuesta de respondientes. Individuos de características similares no tienen por qué ser plenamente coincidentes.
- *Predecir la respuesta a partir de las que da a otras preguntas*. Suele hacerse a través de un análisis de *regresión*, considerando sus respuestas en otras variables.

Por ejemplo, predecir ingresos no declarados a partir de la respuesta que proporcionó en variables relacionadas (estudios, ocupación, edad, género, hábitat u otras que se estimen relevantes).

En todo caso, se trata de respuestas “imputadas” (no “reales”), que serán analizadas como valores observados “reales”, con el riesgo que comporta en la interpretación de los resultados del estudio. Por lo que se aconseja diferenciar los valores “reales” de los “imputados”.

“Los valores imputados pueden ser buenas estimaciones, pero no son datos reales” (Lohr, 2000: 276).

“La mejor imputación es la que no se hace (...). En cualquier caso, si finalmente nos decidiéramos por imputar, siempre es conveniente distinguir entre los casos que tienen valores reales, declarados por el entrevistado, y aquellos que tienen valores imputados” (Sánchez Carrión, 1999: 280).

Se reitera la necesidad de intensificar los esfuerzos para aumentar la tasa de respuesta durante el *trabajo de campo*, antes que aplicar ajustes estadísticos mediante la *ponderación* o *imputación*. Por muy perfecto que sea el ajuste, introduce *sesgos* en las estimaciones de los parámetros poblacionales. Se sobreestiman las respuestas de los “respondientes” en representación de los “no respondientes”, sea o no esta correspondencia cierta.

EJEMPLO DE DISEÑO MUESTRAL Y EJECUCIÓN DEL TRABAJO DE CAMPO

Mejoras introducidas en la *Encuesta Social Europea* (ESE), detalladas por Anna Cuxart y Clara Riba (2009: “Mejorando a partir de la experiencia: La implementación de la tercera ola de la ESE en España”, *REIS* n.º 125: 147-168).

El objetivo principal de la *ESE* es el estudio comparado y longitudinal de las actitudes, valores y comportamientos de los ciudadanos europeos en referencia a los ámbitos económico, social y político. Es una encuesta de carácter bianual. En las tres primeras ediciones participaron un total de 32 países, si bien sólo 17 países (entre ellos España) en todas las ediciones. El *Norwegian Social Science Data Services* (NSD) es el organismo encargado del archivo de las series de datos de cada país, de su integración en un archivo común y de su posterior divulgación. El cuadro 7.30 resume los requisitos técnicos de la ESE establecidos por el equipo de expertos en quienes recae su diseño y ejecución:

CUADRO 7.30. Diseño muestral y trabajo de campo de la Encuesta Social Europea

REQUISITOS TÉCNICOS DE LA ENCUESTA SOCIAL EUROPEA	
POBLACIÓN	– Todos los residentes en domicilios particulares, de edad igual o superior a los 15 años, sea cual sea su nacionalidad, ciudadanía o lengua.
TIPOS DE MUESTREO	– Muestra totalmente <i>probabilística</i> en todas sus etapas de selección, con indicación de las probabilidades de selección de cada una de las unidades muestrales – No se admite ni la aplicación de <i>cuotas</i> ni la <i>sustitución</i> de casos.
TASA DE RESPUESTA	– Objetivo: mínimo del 70%.
TAMAÑO MUESTRAL	– Depende de cada diseño. En todo caso, ha de ser suficiente para garantizar un mínimo de 2.000 entrevistas completadas y un tamaño muestral efectivo ¹ de 1.500.
TRABAJO DE CAMPO	– Redacción de los cuestionarios en todas las lenguas habladas por un mínimo de un 5% de la población. – Entrevista <i>cara a cara</i> . – Período de realización entre uno y cuatro meses. – El primer contacto debe ser personal <i>cara a cara</i> . – Mínimo de <i>cuatro intentos de contacto</i> antes de abandonar una unidad de muestreo por improductiva, de los cuales uno debe ser en horas vespertinas y otro en fin de semana. – Número máximo de entrevistas asignadas a cada entrevistador igual a 48. – Formación específica de entrevistadores con participación del coordinador nacional. – Documentación por parte del entrevistador de todas las incidencias del trabajo de campo en un formulario de contacto. – Compensaciones materiales a los entrevistados.

CUADRO 7.30. (Continuación)

REQUISITOS TÉCNICOS DE LA ENCUESTA SOCIAL EUROPEA

SUPERVISIÓN	<ul style="list-style-type: none"> – Control por parte de la empresa de campo del trabajo de los encuestadores, con supervisión de un 5% de las entrevistas, un 10% de los rechazos y un 10% de los casos no contactados. – Realización de informes quincenales por parte de la empresa que realiza el trabajo de campo, con inclusión de las tasas de respuesta hasta el momento y desglose de negativas claras y negativas suaves. – Asignación de los casos correspondientes a negativas suaves a un segundo entrevistador que debe reiniciar el proceso. – Seguimiento continuo del trabajo de campo por parte del equipo nacional a través de la supervisión de los informes y de los ficheros de datos provisionales suministrados por la empresa encuestadora y envío de resultados al equipo de expertos. – El <i>tamaño efectivo</i> de una <i>muestra</i> es el tamaño de una muestra aleatoria simple con la que se obtendría el mismo grado de precisión.
--------------------	--

¹ El *tamaño efectivo* de una *muestra* es el tamaño de una muestra aleatoria simple con la que se obtendría el mismo grado de precisión.

El cumplimiento de dichos requisitos se toma como garantía de calidad. Respecto a la *muestra*, el diseño en España ha ido evolucionando. En cuanto a la *población*, a partir de la segunda ola aumenta la *cobertura* con la incorporación de las ciudades autónomas de Ceuta y Melilla. La *tasa de respuesta* en Ceuta y Melilla fue del 71%, cifra superior a la del conjunto de la encuesta (55%) y la más elevada de todas las comunidades autónomas.

Otro cambio importante entre la primera y las siguientes olas fue la reducción del número de etapas. En la primera ola se utilizó como *marco de muestreo* el registro de domicilios del INE (Instituto Nacional de Estadística), lo que supuso un diseño muestral en tres etapas. El elevado número de direcciones no válidas (13%) llevó al equipo a reconsiderar la conveniencia de seguir trabajando con domicilios. El propio INE recomendó la utilización como *marco de muestreo* del registro de individuos, que, aunque inicialmente procede del censo, como el de viviendas, es actualizado anualmente a través de los datos de los padrones municipales. De esta manera, las tres etapas, consistentes en selección de una sección, de un domicilio dentro de la sección y de un individuo dentro del domicilio, se reducían a dos. Además, la extracción de una muestra de individuos presentaba la ventaja adicional de la eliminación del sesgo de selección que podía ser introducido en la etapa final de selección aleatoria de un individuo dentro de cada domicilio.

Por lo que respecta a la *estratificación* según tamaño de municipio, la experiencia aconsejó la modificación progresiva de los criterios aplicados. En la primera ola se aplicó uno de los criterios más generalizados: distinguir entre capitales de provincia y resto de municipios. Sin embargo, la existencia de ciudades en España más pobladas que muchas de las capitales de provincia llevó a la inclusión de un nuevo estrato en la segunda ola: el correspondiente a poblaciones de más de 100.000 habitantes que no son capitales de provincia. Los datos mostraron que es en esas grandes ciudades donde se da una menor tasa de respuesta. En la tercera ola se refinó la *estratificación* con cuatro tramos de tamaño de municipio.

Por otra parte, como resultado de la baja *tasa de respuesta* en el País Vasco, en la primera ola (35%), en la segunda se incrementó un 50% el tamaño de la *muestra* correspondiente a esta comunidad. En la tercera ola, dicha sobrerrepresentación sólo se aplicó en los dos tramos superiores de tamaño de municipio, puesto que la participación en pequeñas ciudades y zonas rurales del País Vasco no difería significativamente de las demás comunidades autónomas.

En el cálculo del tamaño de la *muestra* se consideró el llamado *efecto de diseño* (el incremento de la variabilidad de los datos obtenidos por el diseño muestral aplicado en comparación con la variabilidad que se habría obtenido en una muestra aleatoria simple del mismo tamaño). En la segunda ola se estimaron los *efectos de diseño* a partir de los datos de la primera ola. Al utilizar una muestra de individuos, desapareció una de las fuentes de incremento de la variación: la debida a la selección de un individuo dentro de cada domicilio. Por lo que se refiere al efecto de la *conglomeración*, se comprobó que había sido subestimado (coeficiente de correlación intra $\rho = 0,09$ en lugar del 0,02 utilizado en los cálculos iniciales). Se trató de compensar la elevación del valor de ρ con una disminución del número de individuos seleccionados por sección (6 o 7 en vez de 10 o 12) y considerando el efecto de la *estratificación*. Con esta estrategia se consiguió que el *efecto de diseño* total esperado fuera menor que el de la ola anterior. Además, se estimó que con las mejoras introducidas en el *trabajo de campo* se alcanzaría una *tasa de respuesta* cercana al 65% y que el número de casos no válidos se reduciría al 10%, con lo que se obtuvo una reducción considerable en el tamaño bruto de la muestra.

En la tercera ola se optó por seguir las recomendaciones de tomar como base para los cálculos una *tasa de respuesta* media del 62%, que, sin renunciar al objetivo del 70%, era realista. Por otra parte, la calidad de la muestra de individuos usada en la ola anterior sugirió la conveniencia de rebajar hasta el 5% la previsión de casos inválidos. Con ello el número medio de respuestas esperado por sección (3,8) resultó ligeramente inferior al previsto para la muestra de 2004 (4,2), redundando en un menor efecto de diseño debido a la *conglomeración*. En total, el *efecto de diseño* previsto y el tamaño bruto de la *muestra* fueron similares a los de la ola anterior.

En el caso de España, la utilización de CAPI en la segunda y tercera olas ha permitido una mejora notable en la labor de *control* ejercida por el equipo español. Es de destacar el mayor esfuerzo, a partir de la segunda ola, en los intentos de localización y contacto de las unidades muestrales: establecimiento de un mínimo de siete intentos de contacto (en vez de los cuatro requeridos por el equipo de expertos; de ellos dos debían ser en horario vespertino y dos en fin de semana (en lugar de uno y uno establecidos por el equipo de expertos, respectivamente)).

Aunque en la ola más reciente se mantiene el porcentaje de *rechazos*, ha habido una disminución importante de los *casos no localizados*; en parte compensada por un aumento de los casos *inelegibles* (por error en la dirección, cambio de domicilio, demolición de la vivienda, defunción del interesado), principalmente por cambios de domicilio. La estricta *supervisión* del *trabajo de campo* en la tercera ola resultó en la anulación de un puñado de entrevistas por haberse detectado errores en la administración del cuestionario.

Tasa de respuesta global del 66%. En los dos grupos de municipios de menor tamaño (los de menos de 10.000 habitantes y los de 10.001 a 50.000 habitantes) se ha superado el objetivo del 70%. El número de entrevistas necesarias para un tamaño efectivo de 1.500 en la tercera ola fue de 1.904 y el número de entrevistas completadas 1.876 (en la segunda ola fueron 1.829 y 1.663, respectivamente).

Lecturas complementarias

- Blaikie, N. (2010): *Designing social research*. Polity Press. Cambridge.
- Clairin, R. y Brion, P. (2001): *Manual de muestreo*. Hespérides. Madrid.
- Díaz de Rada Igúzquiza, V. (2005): *Manual de trabajo de campo en la encuesta*. Madrid. CIS. Cuaderno Metodológico n.º 36.
- Groves, R. M. et al., (2009): *Survey Methodology*. Wiley & Sons. Hoboken, Nueva Jersey.
- Henry, G. T. (1990): *Practical sampling*. Sage. Newbury Park.
- Kish, L. (1965/1995): *Survey sampling*. John Wiley & Sons. Nueva York.
- Levy, P. S. y Lemeshow, P. S. (1991): *Sampling of populations: methods and applications*. Wiley. Nueva York.
- Lohr, S. L. (2008): “Coverage and sampling”, en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Nueva York: 97-112.
- Rodríguez Osuna, J. (1993): *Métodos de muestreo. Casos prácticos*. Centro de Investigaciones Sociológicas. Colección Cuaderno Metodológico n.º 6. Madrid.
- Sánchez Carrión, J. J. (2000): *La bondad de la encuesta: el caso de la no respuesta*. Alianza. Madrid.
- Singh, S. (2003): *Advanced sampling theory with applications. How Michael selected Amy*. Kluwer Academic Publishers. Dordrecht, The Netherlands.

Ejercicios propuestos

- Quiere analizarse el proceso de adaptación de la población extranjera en el país. Defina la población de estudio. Concrete el *marco* muestral que utilizaría para la selección de la muestra y su repercusión en los errores de encuesta. ¿Habría variaciones dependiendo del método de encuesta que se elija? Justifique la respuesta y diseñe la *muestra*.
- Diseñe la *muestra* para una encuesta a estudiantes universitarios sobre su valoración de la reforma universitaria. Plantee diseños muestrales alternativos en función del método de encuesta elegido y de la población de estudio. Justifique los diferentes pasos hasta la selección de las unidades últimas de observación y su repercusión en *errores de cobertura y muestreo*.
- Si la encuesta se realiza en una universidad que cuenta con 99.610 alumnos matriculados, calcule:
 - El tamaño de la *muestra* para un *error* máximo de $\pm 3,5\%$.

- b) Afije la muestra (en las diferentes opciones posibles), atendiendo al ciclo donde cursa estudios. Describa las diferentes afijaciones señalando las ventajas y límites de cada una.

<i>Ciclo universitario</i>	<i>N.º alumnos</i>	<i>Varianza</i>
Primero	49.309	1.980
Segundo	36.132	2.320
Tercero	14.169	2.418

- c) ¿Qué es ponderar? *Pondere* la muestra para la *afijación simple*.
- d) Si la *muestra* hubiese estado integrada por 978 alumnos, calcule el *error muestral* partiendo de que el 42% de los alumnos es favorable a la reforma universitaria.
4. Se desea conocer el número de horas por término medio que los niños en edad escolar ven la televisión a la semana. Para ello se realiza una encuesta a los padres mediante entrevista personal, obteniéndose que, en los días laborales, sus hijos ven la televisión una media de 1 hora y 20 minutos. La tasa de *no respuesta* fue del 29,7%. Tras el seguimiento telefónico de los no respondientes, la media se eleva a 2 horas y 15 minutos. Calcule el *error de no respuesta* e interprete la disparidad en los resultados.
5. En un estudio sobre discriminación en el ámbito laboral, se desea encuestar a la población ocupada de 16 y más años. Del *censo de población activa* se obtiene la siguiente clasificación en dicho municipio:

Profesión	<i>Población activa ocupada</i>		
	Total	Varones	Mujeres
Profesionales y técnicos	117.430	60.942	56.488
Personal directivo	17.131	15.252	1.879
Empleados administrativos	102.507	49.721	52.786
Comerciantes y vendedores	76.039	41.743	34.296
Personal de servicios	81.409	34.364	47.045
Agricultores	17.377	14.236	3.141
Obreros	292.238	270.952	21.286
Total	704.131	487.210	216.921

- a) Calcular el tamaño de la *muestra* para un *error* máximo de $\pm 2,5\%$ y distribuirla proporcionalmente.

- b) Ponderar la muestra si se hubiesen hecho 400 entrevistas a profesionales y técnicos. También, para la *afijación simple*.
- c) Calcular el *error* para las siguientes estimaciones de insatisfechos con su situación laboral: 39,4% profesionales y técnicos, 28,3% personal directivo, 73,8% empleados administrativos, 54,1% comerciantes y vendedores, 59,6% personal de servicios, 48,2% agricultores y 52,1% obreros.
6. En un municipio de 916.347 habitantes se realiza una encuesta *cara a cara* (en los domicilios de los encuestados) para conocer su grado de satisfacción con la política municipal. La *muestra* la integran 1.372 personas de 18 y más años, de ambos sexos (para un *error* de $\pm 2,7\%$ y máxima heterogeneidad). La tasa de respuesta fue del 68%. Pondere la *muestra* de acuerdo con la población de estudio y la *muestra* original, teniendo en cuenta la información disponible en el cuadro adjunto.

	<i>Población</i>	<i>Muestra inicial</i>	<i>Muestra final</i>	<i>Satisfechos con la política municipal</i>
Varones				
18 - 30	115.407	173	122	73
31 - 50	139.394	209	129	87
51 - 65	101.998	153	95	68
Más de 65	73.448	110	86	65
Mujeres				
18 - 30	121.308	181	139	90
31 - 50	136.407	204	143	104
51 - 65	129.001	193	130	101
Más de 65	99.384	149	89	73
Total	916.347	1.372	933	661

ANEXO: Tabla de números aleatorios

96754	17676	55659	44105	47361	34833	99116	48730	84989	73443	23476	91843
34357	86040	53364	71726	45690	66334	15696	39765	65178	71058	90637	82068
06318	37403	49927	57715	50423	67372	97720	02432	51269	53342	69620	02132
06211	52820	07243	79931	89292	84767	11666	52615	71861	32141	98000	84342
47534	09243	67879	00544	23410	12740	71628	81369	78783	41943	75691	05597
98614	75993	84460	62846	59844	14922	40501	86679	99943	23930	99943	52249
24856	03648	44898	09351	98795	18644	22518	60332	98215	22554	98215	90600
96887	12479	80621	66223	86085	78285	94771	63116	62173	48888	62173	21505
90801	21472	42815	77408	37390	76766	18106	85693	98191	73947	98191	22278
55165	77312	83666	36028	28420	70219	60251	02540	02146	54440	02146	32949
73399	54463	14530	58030	89354	47058	29085	05540	62859	05540	69785	96991
34736	15389	01017	76364	17005	83670	67382	55896	17958	88596	52487	93692
58904	85941	13215	97518	35861	69218	90669	64721	54132	64721	01240	05963
58103	61149	91316	19558	32865	00528	78053	59880	32007	58967	20317	42058
51459	05219	40844	24591	08267	14532	52860	57994	95301	57932	97581	77470
44300	41417	52098	88628	50908	60933	47782	15547	21891	10421	33255	66472
98147	28357	77204	47237	75019	01385	32589	39585	59874	34434	02314	33587
93643	17783	51400	62677	36466	55472	01485	68542	11869	70748	62457	00140
03062	40950	64432	63157	70326	33268	20054	78296	53671	09163	32547	65872
21580	32995	36863	56689	41757	83463	85934	94443	99580	96635	11458	22356
42270	22662	35911	79365	02905	97365	14452	42478	17985	78829	77112	33625
10909	85205	73800	42249	96325	47286	46614	10267	32541	39559	00214	85124
43338	40756	57412	13858	16269	49177	16249	44785	95368	65959	31540	01240
38100	69440	76593	58925	52862	77379	62800	30101	21896	16638	96780	32054
07819	81619	35682	92511	84502	88629	50908	69725	85091	09965	22357	47081
27041	98626	70639	46614	64886	69831	75019	32883	32489	34578	88547	32510
96001	94070	39226	16249	21145	23258	62576	80854	64782	69850	32651	63278
67814	00015	02015	91530	93398	65978	94578	65418	55745	35670	02147	95348
74307	84820	88218	62800	84740	11486	32657	03256	53004	02147	96021	32542
54719	64157	67079	10089	35247	57826	84502	59862	17485	32652	32510	02150
86538	65905	92294	67382	95863	45578	50908	32578	99502	36958	63871	36715
08478	18850	35774	90669	04286	02485	75019	74983	58521	32547	98652	91902
47750	82414	83091	78030	52489	67480	36466	75781	41173	25100	85264	32574
95446	11286	85966	03638	30587	31548	70326	84952	72889	35874	73691	32851
48737	10651	71140	59888	25781	65890	47157	75850	69124	96584	23541	63527
98409	87719	79365	51143	08186	21578	54879	74261	03258	63985	20547	06395
45476	20652	04224	32580	26927	34879	35678	05623	19031	36325	36023	93658
89300	10806	13858	94624	15345	00147	00241	28759	46716	01487	32516	82654
50051	29881	58925	34875	77455	24870	01578	99852	04578	85472	32001	79451
31753	66164	92511	02507	75577	34865	98521	00540	01453	14050	32640	11473
15579	82718	46614	48963	30979	20004	77548	26715	64185	65982	93924	72358
39515	04615	13249	02674	76305	05241	03251	65988	66782	54213	17918	01145
68624	25670	91530	13570	75779	62481	42579	27581	23157	33258	95328	87524
78295	16706	10089	25796	90279	65891	53262	32689	12248	26054	21853	02150
94443	55368	62800	55820	37231	96488	14218	99580	02358	65849	63795	45802

8

ANÁLISIS DE DATOS Y PRESENTACIÓN DE RESULTADOS

A la recogida de información sigue la fase, también clave, de análisis de datos. Una fase muy beneficiada por los progresivos avances y accesibilidad de los paquetes estadísticos, aunque no la eximen de errores. Errores que pueden llegar a nublar los esfuerzos realizados en estadios precedentes de la investigación. Unos pueden ser propiciados por actuaciones erróneas en la fase previa de preparación de los datos para el análisis (edición, codificación, grabación), otros por la inadecuación de la técnica estadística aplicada a los datos que se analizan o por erróneas interpretaciones de los resultados de los análisis, nublando los esfuerzos de fases anteriores de la investigación.

A lo largo de este capítulo se va a ofrecer una breve panorámica de técnicas estadísticas de carácter univariable, bivariable y multivariable, dependiendo del número de variables que participen del análisis. Los condicionantes de espacio, unido al amplio abanico de técnicas existentes, hacen que la exposición se limite a meros trazos descriptivos de características definitorias. Para un conocimiento más pormenorizado se remite a manuales específicos de análisis de datos (algunos citados al final del capítulo). El capítulo concluye con la redacción del *informe* de la investigación.

8.1. Preparación de los datos para el análisis

Antes de comenzar los análisis estadísticos, han de cumplimentarse varias actuaciones previas que posibiliten su realización. Incluye la *edición* (o primera inspección de los datos) y su *grabación* (cuando los datos no se han recogido mediante modos informatizados). También la *codificación* (de preguntas *abiertas* en un cuestionario), el cálculo de pesos o *coeficientes de ponderación* (para equilibrar la muestra con respecto a la población de estudio, por la existencia de *errores de no observación*: no cobertura, no respuesta y de muestreo), la *imputación* (cuando ésta se elija como estrategia de corrección de los *errores de no respuesta*) y la *inspección explo-*

ratoria de los datos grabados para su posterior *tabulación*, que dará o no pie a análisis estadísticos más complejos.

En cada una de estas actuaciones pueden producirse “errores”: de *codificación* (que no se ajuste al procedimiento prescrito), *grabación* de los datos (introducción de códigos de respuesta incorrectos), de asignación de *pesos*, de *tabulación*. Errores que no siempre son fácilmente identificables y que pueden traducirse en errores *sistemáticos*, mermando la calidad de la investigación.

“Un investigador que tenga una muestra perfecta, medidas perfectas, y ningún error en la recogida de datos, pero que cometa errores en el proceso de codificación o al grabar los datos en el ordenador, puede arruinar un proyecto de investigación entero” (Neuman, 2000: 316).

- La fase de *edición* incluye diferentes actuaciones dirigidas a detectar datos registrados erróneamente para su corrección antes de proceder a su análisis.

Por ejemplo, en la investigación mediante encuesta, los supervisores o el personal responsable del *trabajo de campo* han de inspeccionar los cuestionarios antes de darlos por válidos y remitirlos al personal responsable de su grabación informática (en caso de cuestionarios de lápiz y papel). Preguntas sin respuesta y respuestas no claramente especificadas o ambiguas han de ser resueltas para que no queden descartadas de los análisis estadísticos (remitiéndoselas a los entrevistadores, contactando con los propios encuestados...). Por ello es importante que los cuestionarios incluyan un apartado que identifique al entrevistador y al encuestado. La detección de errores graves que no puedan solventarse lleva a excluir cuestionarios de los análisis.

- La *codificación* implica poner los datos brutos en un formato que sea legible por el programa estadístico que se utilice. Como se explicó en el capítulo 6, la *codificación* de los cuestionarios suele realizarse antes de la *grabación* de los cuestionarios. Si bien, la *recodificación* de las respuestas a preguntas *abiertas* o *cerradas* se posterga a la fase de preparación o *depuración de los datos* para el análisis.

Por ejemplo, la *codificación* de la pregunta *abierta* relativa a la ocupación laboral, que en las encuestas que realiza el Centro de Investigaciones Sociológicas (CIS) adopta el siguiente formato de pregunta:

“¿Cuál es/era su actual/última ocupación u oficio? Es decir, ¿en qué consiste/tía específicamente su trabajo? (Precisar lo más posible las actividades realizadas, EJEMPLO: mecánico, reparador de automóviles, ayudante de odontología, profesor de enseñanza primaria, etc.). Nos referimos a su *ocupación principal*: aquella por la que Ud. (o el cabeza de familia) obtiene/nía mayores ingresos.

Para su *codificación* el CIS aplica los códigos CNO (Clasificación Nacional de Ocupaciones) de 1979. Éstos van desde los códigos 011 (Químicos) hasta 999 (Trabajadores no clasificados en otros grupos primarios: peones). El primer dígito corresponde a grandes grupos ocupacionales (0/1 Profesio-

nales, técnicos y similares, 2 Miembros y personal directivo de órganos de la Administración Pública y directores y gerentes de empresas, 3 Personal de servicios administrativos y similares, etc.). El segundo y el tercer dígitos a categorías específicas de cada grupo. Los códigos asignados podrán después *re-codificarse* en un número inferior de categorías (de acuerdo con las frecuencias de respuestas), en la fase posterior de análisis estadísticos.

- En la *codificación* puede haber errores porque se introducen códigos de respuesta incorrectos (por despiste, mala formación...); o por deficiencias en la codificación diseñada, que no abarca todas las respuestas de la pregunta *abierta* (confusión de qué códigos aplicar a respuestas no claramente especificadas en el *libro de códigos*, y pérdida consiguiente de uniformidad en la *codificación* de todos los cuestionarios, cuando son varias las personas que en ella participan). Biemer y Lyberg (2003) estiman que los *errores de codificación* pueden suponer el 20% de todos los errores de encuesta.
- Tras la grabación de las respuestas en la *matriz de datos* (con el código numérico asignado a ellas) prosigue la revisión de las respuestas y posterior *depuración de los datos*: se comprueba si los códigos están comprendidos en el rango de valores posibles.

Por ejemplo, en la variable “sexo” (1 varón, 2 mujer, 9 no contesta), encontrar códigos de respuesta no incluidos (3, 4, 8 u otro). Supondrá la revisión del cuestionario en busca del código correcto.

Pero si el error se debe a un mal registro de la respuesta en el cuestionario, su corrección es más compleja. Supondrá ir a la fuente originaria de la información. Por otra parte de detección de errores se complica, cuando el código de respuesta grabado erróneamente sí está incluido en el rango de valores posibles de la pregunta.

Por ejemplo, en la variable edad anotar 54 años en vez de 45.

- En casos con muchas preguntas *sin respuesta* (“*missing values*”) puede optarse por su exclusión de los análisis, excepto que al investigador le interese su inclusión. Las preguntas con un elevado porcentaje de *no respuesta* puede también descartarse de los análisis. Pero otra opción es corregirlos aplicando coeficientes de *ponderación e imputación*, como se explicó en el capítulo 7. Decisiones también presentes en la fase de preparación de los datos para el análisis.

Para cualquiera de las actuaciones que incluye la *depuración de los datos* son de gran ayuda los análisis *univariantes exploratorios* (observar la distribución de frecuencias absolutas y relativas, la *no respuesta*, estadísticos univariantes, representaciones gráficas) y de cruce de variables. Algunos *errores de grabación* pueden detectarse cruzando respuestas de preguntas relacionadas. *Por ejemplo*, que la edad de la madre sea inferior a la del hijo o que la persona figure como no usuaria de Internet y responde preguntas relativas al uso de Internet.

EJEMPLO DE GRABACIÓN DE CUESTIONARIOS

Los *ficheros de datos* del Centro de Investigaciones Sociológicas se hallan en el formato tradicional *ASCII* (*American Standard Code for Information Interchange*), para que puedan aplicarse en cualquier programa estadístico (SPSS, BMPD, SAS). Cada *fila* representa un cuestionario (los casos observados); las *columnas* incluyen los códigos de respuestas de cada pregunta del cuestionario. Cada cuestionario se considera un registro, cuya longitud depende del número de preguntas que incluya (traducidas a variables). En el cuestionario, si está precolumnado, se especifica la posición que cada variable (pregunta) ocupa en el *fichero de datos*. Al fichero de datos se suma el de *sintaxis* o de *definición de datos*. Éste comprende las instrucciones precisas para la lectura del *fichero de datos*. Las instrucciones incluyen información primordial de las variables en el cuestionario: las *columnas* donde están ubicadas en el *fichero de datos*, las *etiquetas* de cada variable y sus atributos (categorías de respuesta o valores numéricos) y la especificación de la codificación de la *no respuesta*, en cada pregunta, para su tratamiento estadístico.

El cuadro 8.1 extrae unas líneas del *fichero de datos* de la encuesta CIS-OBERA-XE de *Actitudes ante la inmigración* de 2010 (estudio CIS n.º 2.846). El cuadro 8.2, el correspondiente *fichero de definición de datos*. El cuadro 8.3, su transformación en *fichero de datos* en versión SPSS.sav (para Windows). Esta última conjuga ambos ficheros (variables en columnas, casos en filas).

CUADRO 8.1. Fichero de datos

```
284600622 111 031 0 0 01 1 899 19999334298 21120 0 5 03 3 3 5 0 0 0 7 9321 6 6 6 61511111144244144...
284600623 111 031 0 0 01 1 999 199993333 4 1 920 010 02 7 0 5 0 0 0 5 4838 5 5 5 58411111144418141...
284600624 111 031 0 0 01 8 199 199992132 6161525 010 0398 8 8 1 1 1 1 9883 6 7 5 5611211113141414...
284600625 111 031 0 0 01 1 899 199993333 4 1 920 010 02 7 0 5 0 0 0 5 4838 5 5 5 584111111444181413...
284600626 111 031 0 0 01 8 199 199992132 6161525 010 0398 8 8 1 1 1 1 9883 6 7 5 5611211113141412...
```

CUADRO 8.2. Fichero de definición de datos

```
DATA LIST FILE='DA2846.dat'
  /CUES 5-9 A1 1-9(A) CCAA PROV 10-13 MUN 14-16 TAMUNI 17 EIEX 18
  DISTR 19-20 SECCION 21-23 ENTREV 24-27
  P0 28 P101 TO P103 P201 TO P203 29-40 P301 TO P304 41-44
  P4 P501 p502 p6 p6a p7 p7a 45-58 p8 59 P901 TO P908 60-75
  P10 TO P12 76-78 P1301 TO P1304 79-86 P14 P15 P1601 TO P1606
  P1701 TO P1705 P1801 TO P1805 P19 87-105
  P20 P21 106-109 P2201 P2202 P23 P24 P2501 TO P2505 P2601 TO P2604
  P27 P2801 TO P2808 P2901 TO P2904 110-135 P30 P31 P31A 136-141
  P32 TO P35 142-145 P36 p36A01 TO P36A10 P37 P37A P38 P38A 146-160
  P39 P40 161-164 P41 165 P41A 166-167 P42 P42A P43 P44 168-171
  P45 172-173 P46 P47 174-175 P47A 176-177 P48 P49 178-179
  P49A 180-181 P50 P50A P51 TO P55 182-188 P56 189-191
  P57 P57A 192-193 P58 P59 194-197 P60 198 P60A 199-201
```

[.../...]

CUADRO 8.2. (Continuación)

P61 P6201 TO P6204 202-206 P6301 TO P6305 207-221(A)
P64 TO P67 222-225 FINAL 226-273(A) PESO 274-278.

COMPUTE PESO=PESO/10000.
WEIGHT BY PESO.

MISSING VALUES P0 TO P8 P10 TO P12 P14 TO P38A P41 TO P47 P48 TO P6204 P64 TO P67 (0).

VARIABLE LABEL

CCAA 'Comunidad autónoma'
/PROV 'Provincia'
/TAMUNI 'Tamaño de hábitat'
/MUN 'Municipio'
/DISTR 'Distrito'
/P301 'Personas mayores que viven solas'
/P302 'Pensionistas'
/P303 'Parados'
/P304 'Inmigrantes'
/P901 'Que tenga un buen nivel educativo'
/P902 'Que tenga familiares cercanos viviendo aquí'...

VALUE LABEL

CCAA
01 'Andalucía' 02 'Aragón' 03 'Asturias (Principado de)'
04 'Balears (Islas)' 05 'Canarias' 06 'Cantabria'
07 'Castilla-La Mancha' 08 'Castilla y León'
09 'Cataluña' 10 'Comunidad Valenciana' 11 'Extremadura'
12 'Galicia' 13 'Madrid (Comunidad de)' 14 'Murcia (Región de)'
15 'Navarra (Comunidad Foral de)' 16 'País Vasco' 17 'Rioja (La)'
/PROV
01 'Álava' 02 'Albacete' 03 'Alicante' 04 'Almería' 33 'Asturias'
05 'Ávila' 06 'Badajoz' 07 'Balears' 08 'Barcelona' 09 'Burgos'...

FREQUENCIES VAR=CCAA To P67.

CUADRO 8.3. Fichero de datos SPSS.sav

	cues	a1	ccaa	prov	mun	tamuni
1	622	284600622	1	11	0	3
2	623	284600623	1	11	0	3
3	624	284600624	1	11	0	3
4	625	284600625	1	11	0	3
5	626	284600626	1	11	0	3

Además pueden crearse *subficheros* específicos de datos para facilitar los análisis, cuando el tamaño de la muestra es elevado y el fichero de datos adquiere un elevado volumen. Los *subficheros* incluyen sólo las *variables* de interés para la ejecución de análisis concretos o una selección de *casos*.

Por ejemplo, dividir la muestra al azar en dos submuestras (de *análisis* y *validación*) con el propósito de replicar los datos y comprobar si se alcanzan los mismos resultados (dar *validez* a los análisis realizados). Si el tamaño de la *muestra* es elevado, puede destinarse cada mitad a muestra de análisis y de validación); si es pequeña, el peso de la muestra de análisis suele ser mayor (60% e incluso 70%) que la de validación.

8.2. Análisis exploratorio univariable

Los análisis *exploratorios* ayudan a detectar *errores* de grabación y de codificación, además de proporcionar la primera aproximación a los datos a analizar. De la exploración puede deducirse qué tratamiento dar a la *no respuesta*, si se *recodifican* variables (agrupando o modificando las categorías de respuesta), o si se crean nuevas variables combinando dos o más variables, aparte de decidir los siguientes análisis estadísticos a realizar.

Primero se procede al análisis, por separado, de cada variable mediante *tablas de frecuencia*, *gráficos* y *estadísticos univariados* (que resumen la distribución de la variable en la *muestra*). La secuencia analítica sigue con la elaboración de *tablas de contingencia* en las que se cruzan dos, tres e incluso cuatro variables, que actúan como variables de *control* en la relación entre dos variables. La mayoría de los informes se limitan a la interpretación de frecuencias porcentuales *univariados* y de *tablas de contingencia*. Si bien, cada vez más los análisis prosiguen con la aplicación de técnicas analíticas *multivariados* para analizar las interrelaciones entre más de dos variables.

La *tabla de frecuencias* incluye los valores de la variable (distribuidos en clases o categorías), junto con su frecuencia *absoluta* (número de casos) y *relativa* (porcentaje que representa en el conjunto de la *muestra*). Permite detectar errores de grabación (como ya se ha indicado), además de conocer la distribución de la variable, la *no respuesta de ítem* y la conveniencia de *recodificar* la variable para análisis posteriores (agrupando valores en un número inferior de categorías o cambiando sus códigos numéricos para hacerlos acordes a las categorías que designan).

En variables *continuas* con un número elevado de valores, como la variable *edad* cuando se graba la edad exacta (“¿Cuántos años cumplió Ud. en su último cumpleaños?”), la *tabla de frecuencia* precisa la previa agrupación de los valores numéricos en intervalos amplios para que adquiera una menor dimensión (en grandes grupos de edad). Bryman y Cramer (1995) recomiendan que el número de categorías esté comprendido entre 6 y 20. Argumentan que menos de 6 y más de 20 categorías puede distorsionar la *forma* de la distribución de la variable.

La información incluida en la *tabla de frecuencias* puede representarse de manera gráfica. Las posibilidades gráficas son numerosas y de fácil realización, al incluirse no sólo en los paquetes estadísticos especializados, sino también en los procesadores de texto de uso común (como el programa *Word*, de Microsoft), que amplían el abanico de opciones gráficas inclusive en tres dimensiones. Entre las opciones gráficas tradicionales están:

- *Diagrama de barras*: una serie de “barras”, para cada categoría de la variable, cuyas longitudes dependen de su frecuencia.
- *Histograma*: sustituye las barras por rectángulos, adecuándose más a variables continuas.
- *Gráfico de sectores*: como una “tarta” dividida en porciones (“sectores”) de diferentes trazas o colores (para distinguir unos de otros). Sus ángulos dependen del porcentaje de casos en cada categoría de la variable.
- *Polígono*: un gráfico lineal que une los valores de la variable y cuya altura es proporcional a su frecuencia.
- *Ojivas*: polígonos de frecuencias acumuladas. Muestran la frecuencia de casos por encima, o por debajo, de un determinado valor de la distribución.
- *Tallo y hojas*: un gráfico similar al histograma, pero integrado por los dígitos de los valores de las variables. Suele tomarse como referente de las medidas de tendencia central de una distribución de frecuencias. El dígito a la izquierda (el *tronco*) que comprenda más valores a la derecha (las *hojas*) será aquél que agrupe un mayor número de casos.
- “*Caja*”: ofrece una visión global y sintética de la distribución de la variable. El 50% de los casos quedan comprendidos en el interior de la “caja”. En los extremos se sitúan, respectivamente, el 25% superior e inferior. Con un asterisco se señala la mediana, con una X los valores máximos y mínimos y con una O los atípicos (“*outliers*”).

Una información más precisa de la distribución de la variable se obtiene a través de *estadísticos univariados*, que miden la concentración de valores y su representatividad. Incluyen:

Medidas de tendencia central: describen cómo se agrupan los valores de cada variable.

- La *media aritmética* es la más representativa, pero exige que la variable sea continua. Se obtiene de la suma de todos los productos de cada valor por su frecuencia dividida por el número total de casos. Su valor expresa el promedio de los valores de la distribución. Su correcta interpretación precisa del cálculo de medidas de *dispersión*, que gradúen su representatividad, al estar muy afectada por los valores extremos de la variable.
- La *mediana* es el segundo estadístico en importancia. Proporciona el valor que divide a la distribución en dos partes iguales. Permite conocer los valo-

res de mayor y menor presencia en la *muestra*. Su cálculo exige que la variable sea, al menos, *ordinal* (que sus atributos estén ordenados de más a menos o de menos a más; caso de las variables clase social, nivel de estudios, grado de satisfacción...).

- La *moda* puede calcularse con cualquier tipo de variable, al ser el nivel de medición mínimo exigido el *nominal*. Indica el valor de mayor frecuencia en la distribución, aquel que más casos comparten. Pero, puede haber más de una *moda* (distribuciones *multimodales*), lo que dificulta su interpretación.
- Los *cuantiles* representan los valores que dividen a la distribución en partes iguales. Incluyen los *cuartiles* (la divide en cuatro partes iguales), los *deciles* (en diez partes) y los *percentiles* (en cien partes). Su cálculo se asemeja al de la *mediana*. De hecho, el *cuartil* dos, el *decil* cinco y el *percentil* cincuenta expresan el valor mediano.

Medidas de dispersión: permiten conocer cómo varían los valores en torno a la *media* o la *mediana*. Las medidas de dispersión *absolutas* más comunes son el *rango* (o recorrido), la *desviación típica* y la *varianza*.

- El *rango* expresa el número de valores que incluye la variable. Se obtiene de la diferencia entre el valor superior y el inferior. Su comprensión es sencilla, aunque presenta el inconveniente de ser susceptible a la distorsión proporcionada por valores muy extremos en una distribución. Se debe a que sólo considera el primer y el último valor de la variable. Cuando la variable es ordinal o métrica (de intervalo o de razón) pueden calcularse rangos que informan de la distribución de los valores centrales de la variable: *rango intercuartílico* (diferencia entre el tercer cuartil y el primero, incluyendo al 50% de los valores centrales de la distribución), *semiintercuartílico* (el 25% de los valores centrales), *interpercentílico* (el 80% de los valores centrales, al ser la diferencia entre el percentil 90 y 10), y el *semiinterpercentílico* (el 40%).
- La *desviación típica* es el promedio de la desviación de los casos con respecto a la *media*. Como indicador de heterogeneidad, su cálculo se exige para medir la representatividad de la *media*. Al igual que ésta, su valor viene expresado en la unidad de medición de la variable. Para su correcta interpretación se precisa calcular alguna medida de dispersión relativa, como el *coeficiente de variabilidad de Pearson*. Una medida estandarizada que se obtiene dividiendo la *desviación típica* entre la *media aritmética*. De esta forma se obtiene la proporción de heterogeneidad de la variable, siendo de gran utilidad en la comparación de dos o más grupos respecto a una o más variables. Esto únicamente es posible a través de alguna medida estandarizada que permita comparar variables en diferentes unidades de medición (*por ejemplo*, ingresos medidos en euros y en dólares; o comparar la representatividad de las medias de las variables ingresos y edad). Lo mismo puede hacerse con la *mediana*, mediante el *coeficiente de variabilidad de la me-*

diana de Pearson, que se obtiene dividiendo la desviación media de la mediana entre la mediana.

- La *varianza* es el cuadrado de la *desviación típica*. Expresa la heterogeneidad de la variable en los casos analizados. Sus características se asemejan a la *desviación típica*.

Medidas de la forma de la distribución: comprenden la disposición espacial de los valores de una distribución, su *asimetría* y *curtosis*.

- La *asimetría* expresa cómo se agrupan las frecuencias en la curva de la distribución de la variable. Un valor de cero significa que la curva es *simétrica* (el área es la misma a ambos lados de la curva, al coincidir la media con la mediana y la moda). Un valor mayor de cero denota que la distribución es *asimétrica a la derecha* o con sesgo positivo (los casos se agrupan a la izquierda de la curva, debido a una mayor presencia de valores por debajo de la *media*). Un valor inferior a cero corresponde a una distribución *asimétrica a la izquierda* o con sesgo negativo (la concentración de valores se da a la derecha de la curva por la mayor presencia de valores superiores a la *media*). A partir de 0,8 (con signo positivo o negativo) se considera que la *asimetría* es importante.
- La *curtosis* indica la concentración de valores en torno a la *media*. Cuando existe una elevada concentración, siendo la forma de la distribución muy apuntada y elevada, se está ante una distribución *leptocúrtica*. Por el contrario, cuando la forma es muy aplanada, debido a la elevada dispersión de valores en torno a la *media*, se está ante una distribución *platicúrtica*. Ésta coincide con *medias* no representativas. Un valor de *curtosis* igual a 3 significa que la distribución es normal o *mesocúrtica*; un valor mayor de 3, *leptocúrtica*; y menor de 3, *platicúrtica*.

El cálculo del *error típico de estimación* de cualquiera de estos estadísticos permite conocer su significatividad estadística: en qué medida el valor estimado en la *muestra* es inferible a la *población* de la que se extrae. Para ello, se divide el estadístico estimado (media, mediana, coeficiente de asimetría, de curtosis...) entre el *error típico* (que expresa la oscilación en el valor del estadístico de una *muestra* a otra de la misma *población*). De este cociente resulta un valor *t* empírico que, para que sea estadísticamente significativo, ha de ser mayor o igual a 1,96 (con signo positivo o negativo), para el nivel de significatividad usual de 0,05, si la muestra supera los 100 casos. La significatividad indica la probabilidad de *error* en la *inferencia estadística*. Para muestras inferiores, el valor de referencia varía en función de los grados de libertad (igual al tamaño muestral menos uno).

EJEMPLO DE EXPLORACIÓN UNIVARIABLE

Los siguientes datos corresponden a la encuesta CIS-OBERAXE de *Actitudes hacia la inmigración IV*, de septiembre-octubre de 2010 (estudio n.º 2.846), analizada en M.^a Á. Cea D’Ancona y M. S. Valles (2011: *Evaluación del racismo y la xenofobia en España. [Informe 2011]*. Ministerio de Trabajo e Inmigración. OBERAXE).

Primero se extrae la *tabla de frecuencias* como figura en la salida del programa SPSS, y que corresponde a la pregunta número 8 del cuestionario. La pregunta se enuncia en los términos siguientes: “*En su opinión, ¿el número de inmigrantes que hay actualmente en España es insuficiente, aceptable, elevado o excesivo?*”. La tabla incluye la *frecuencia absoluta* (número de casos en cada valor de la variable “percepción de la presencia de inmigrantes”), la *frecuencia relativa porcentual* respecto al total de la muestra (2.800 casos), la porcentual *válida* (una vez eliminados los casos sin respuesta: “no sabe / no contesta”), y el *porcentaje válido acumulado* (que expresa el porcentaje de casos que están hasta un determinado valor de la distribución). Por ejemplo, el 46% de los 2.800 encuestados respondieron “excesivo”. Representan el 48% de los 2.694 que respondieron a la pregunta. Las otras respuestas aglutinan el 52% de las respuestas pronunciadas.

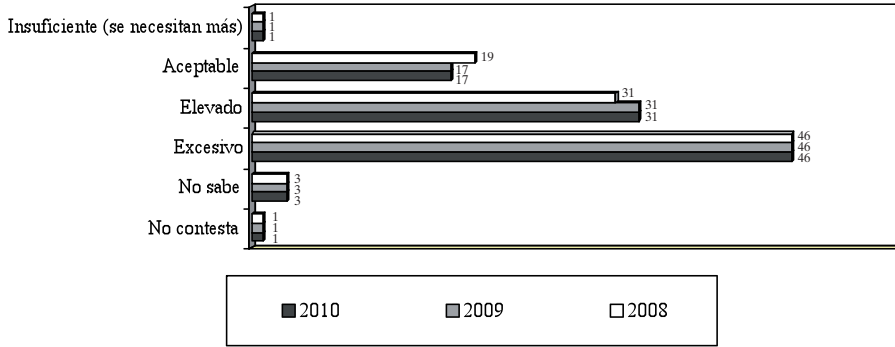
Número de inmigrantes

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Insuficiente (se necesitan más)	11	,4	,4	,4
	Aceptable	466	16,6	17,3	17,7
	Elevado	936	33,4	34,7	52,4
	Excesivo	1282	45,8	47,6	100,0
	Total	2694	96,2	100,0	
Perdidos	Sistema	106	3,8		
Total		2800	100,0		

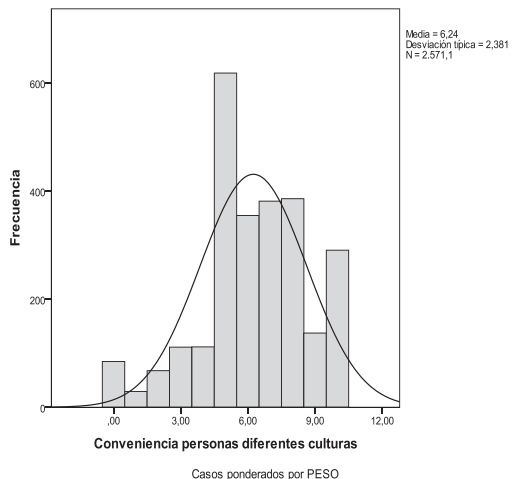
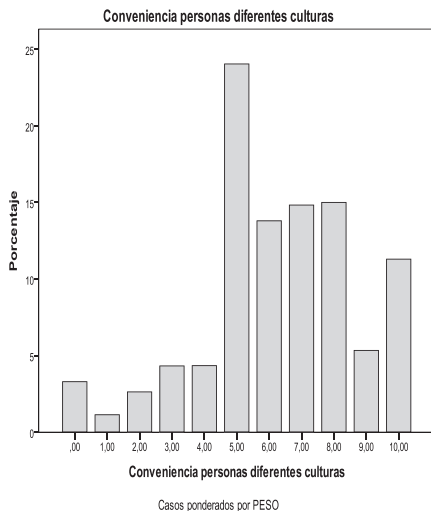
En el *Informe 2011* esta pregunta se incluye en el capítulo 1, dedicado a la *política y presencia inmigratoria*. El abanico de respuesta se desglosa en un gráfico de barras, que incorpora las otras dos encuestas precedentes que incluyen la misma pregunta. De esta forma puede analizarse la evolución de las distintas opciones de respuesta (incluida la *no respuesta*), en términos de frecuencia relativa porcentual (que posibilita el análisis comparativo de respuestas con diferentes bases muestrales).

Percepción de la presencia de inmigrantes en España.
Encuestas CIS-OBEXAXE de 2008 a 2010 (%)

“En su opinión, ¿el número de inmigrantes que hay actualmente en España es...?”



El gráfico anterior se ha realizado con el programa *Word*. En el SPSS, el análisis de *frecuencias* ofrece las opciones de gráficos de barras, de sectores e histograma (con o sin curva normal), tanto para frecuencias como para porcentajes. Para ilustrarlo, se ofrecen los gráficos correspondientes a la pregunta 13: “En una escala de 0 a 10, donde 0 significa “muy negativo” y 10 “muy positivo”, valore el hecho de que la sociedad española esté compuesta por personas de diferentes culturas”. El primer gráfico es el gráfico de barras y el segundo el *histograma* con curva normal, que permite visualizar el distanciamiento de la distribución normal.



El distanciamiento respecto de la distribución normal puede graduarse mejor mediante los estadísticos descriptivos que incluye la tabla adjunta. Se está ante una distribución ligeramente leptocúrtica y asimétrica a la izquierda. Hay una mayor concentración de casos por encima de la *media*. La valoración *media* es de 6,24 (con un *error* de estimación de 0,047) y la *mediana* es 6,00 (la mitad de los casos han dado una respuesta igual o inferior a este valor, la otra mitad un valor superior); la *moda* o valor que concentra un mayor porcentaje de respuesta ha sido el 5,00, como puede verse tanto en el gráfico de *barras* como en el *histograma*.

Descriptivos

			Estadístico	Error típico
Conveniencia personas diferentes culturas	Media		6,2442	,04695
	Intervalo de confianza para la media al 95%	Límite inferior	6,1522	
		Límite superior	6,3363	
	Media recortada al 5%		6,3570	
	Mediana		6,0000	
	Varianza		5,668	
	Desviación típica		2,38085	
	Mínimo		,00	
	Máximo		10,00	
	Rango		10,00	
	Amplitud intercuartil		3,00	
	Asimetría		-,453	,048
	Curtosis		,135	,097

8.3. Análisis cruzado bivariable

El análisis cruzado de dos o más variables es de mayor interés a efectos analíticos. La mayoría de los informes convencionales suelen reducirse a la interpretación de tablas en las que se cruzan los valores de dos, tres o cuatro variables (cuando se introducen variables de control). La presentación de las *tablas* sigue una serie de convenciones que ayudan a su interpretación. Las pautas a seguir de uso común:

- Cuando se analiza la relación causal entre dos variables, la *dependiente* (cuya variabilidad trata de analizarse) suele situarse en las *filas*, mientras que la *independiente* (o predictora) se ubica en las *columnas*. Se procede a la inversa cuando el número de categorías o valores de la variable *independiente* sea elevado (y precise un mayor espacio). En este caso se ubicaría a la inversa: la variable *dependiente* en *columnas* y la *independiente* en *filas*.

- Que la variable *independiente* esté en columna facilita la lectura comparativa de los porcentajes de respuesta en cada valor de la variable *dependiente* para cada valor o categoría de la *independiente* (lectura en horizontal). Como los *porcentajes* se calculan respecto a la variable *independiente*, con objeto de medir su incidencia en la *dependiente*, los porcentajes serán *verticales* (al estar la variable independiente ubicada en columnas) y las comparaciones entre ellos se harán en sentido horizontal. En cada categoría de la variable *dependiente* se procede a la comparación de los porcentajes habidos en cada categoría de la *independiente*. Cuando se prefiere poner la variable *independiente* en las filas y la *dependiente* en columnas, los porcentajes serán horizontales y las comparaciones se harán en sentido vertical. Las comparaciones porcentuales siempre se realizan en la dirección opuesta de la que se calcularon los porcentajes.
- Para que las diferencias *porcentuales* sean relevantes, han de exceder los cinco puntos porcentuales. Más cuando las *bases muestrales* sobre las que se calculan los porcentajes son pequeñas, con el consiguiente incremento del *error típico de estimación* (y su repercusión en la pérdida de *significatividad* estadística de las diferencias porcentuales muestrales). Razón por la cual se recomienda que en la *tabla* aparezca la *base muestral* sobre la que se calculan los *porcentajes*, con preferencia al total de 100%, que se sabe han de sumar los porcentajes. De esta forma podrá mejor calibrarse la magnitud de las diferencias porcentuales.
- En la *tabla* o cuadro ha de señalarse (de manera explícita) la *dirección* en la que están calculados los *porcentajes*: horizontales o verticales. Cuando la tabla no incluye todas las categorías de las variables, ha de indicarse que son porcentajes de cada combinación (y que, en consecuencia, no suman 100). Siempre que se trate de datos muestrales, la precisión de poner los *porcentajes* con decimales parece innecesaria porque toda estimación muestral está sujeta a *error muestral* y demás errores de *no observación* y *medición* descritos en el capítulo 5.
- A la *tabla* ha de adjuntarse un *título* que sintetice su contenido: las variables que comprende y la relación entre ellas.
- Para facilitar la lectura e interpretación de la *tabla*, pueden destacarse las diferencias porcentuales (entre valores o categorías de las variables) de mayor significatividad. Existe relación o asociación entre las variables incluidas en una *tabla*, cuando hay diferencias porcentuales entre las diferentes categorías de la variable *independiente* (para cada valor de la *dependiente*), siguiendo los porcentajes una pauta o trayectoria sin altibajos. Que los porcentajes no varíen de unas celdillas de la tabla a otra significa inexistencia de relación entre las variables.

EJEMPLO DE TABLA DE VARIABLES CRUZADAS

Desea comprobarse si la *edad* del encuestado afecta a su *percepción del número de inmigrantes*. Para ello se cruza la variable antes descrita con la *edad*. Se parte de la hipótesis antes probada de que la percepción en exceso de la presencia de inmigrantes aumenta con la *edad* (del autóctono). En el cuestionario, la pregunta relativa a la *edad* del encuestado adopta un formato abierto, con el siguiente enunciado de pregunta: “¿Cuántos años cumplió Ud. en su último cumpleaños?”. Es una variable *métrica y continua*. Para la realización de la *tabla*, se procede antes a agrupar las edades en intervalos. Se opta por su síntesis en cuatro intervalos: “menos de 30” (1), “30-49” (2), “50-64” (3), “65 y más” (4). La *percepción del número de inmigrantes* (“En su opinión, el número de inmigrantes que hay actualmente en España es...?”) en cambio conserva las mismas cuatro categorías de respuesta que la pregunta presenta en el cuestionario: “Insuficiente (se necesitan más)” (1), “Aceptable” (2), “Elevado” (3), “Excesivo” (4).

La *tabla* incluye la frecuencia absoluta o *recuento* (número de casos) en cada casilla de la tabla (que resulta del cruce de cada *columna* con cada *fila* de la *tabla*); la *frecuencia esperada* (la que cabría esperar en caso de inexistencia de relación entre las variables. Se precisa para el cálculo del estadístico *chi-cuadrado*); el *residuo* no tipificado (la diferencia entre la frecuencia *observada* y la *esperada*) y los *porcentajes* calculados respecto a cada variable (por *filas* y *columnas*).

Tabla de contingencia Número de inmigrantes* Edad (tablas)

			Edad (tablas)				
			1,00	2,00	3,00	4,00	Total
Número de inmigrantes	1,00	Recuento	1	5	2	3	11
		Frecuencia esperada	2,0	4,1 2,5	2,4	11,0	
		% dentro de Número de inmigrantes	9,1%	45,5%	18,2%	27,3%	100,0%
		% dentro de Edad (tablas)	,2%	,5%	,3%	,5%	,4%
		Residuo	-1,0	,9	-,5	,6	
	2,00	Recuento	104	218	88	57	467
		Frecuencia esperada	85,6	172,6	105,2	103,6	467,0
		% dentro de Número de inmigrantes	22,3%	46,7%	18,8%	12,2%	100,0%
		% dentro de Edad (tablas)	21,1%	21,9%	14,5%	9,5%	17,3%
		Residuo	18,4	45,4	-17,2	-46,6	
	3,00	Recuento	197	349	203	187	936
		Frecuencia esperada	171,6	345,9	210,8	207,7	936,0

[.../...]

		Edad (tablas)				Total
		1,00	2,00	3,00	4,00	
	% dentro de Número de inmigrantes	21,0%	37,3%	21,7%	20,0%	100,0%
	% dentro de Edad (tablas)	39,9%	35,0%	33,4%	31,3%	34,7%
	Residuo	25,4	3,1	-7,8	-20,7	
4,00	Recuento	192	424	314	351	1281
	Frecuencia esperada	234,8	473,4	288,5	284,2	1281,0
	% dentro de Número de inmigrantes	15,0%	33,1%	24,5%	27,4%	100,0%
	% dentro de Edad (tablas)	38,9%	42,6%	51,7%	58,7%	47,5%
	Residuo	-42,8	-49,4	25,5	66,8	
Total	Recuento	494	996	607	598	2695
	Frecuencia esperada	494,0	996,0	607,0	598,0	2695,0
	% dentro de Número de inmigrantes	18,3%	37,0%	22,5%	22,2%	100,0%
	% dentro de Edad (tablas)	100,0%	100,0%	100,0%	100,0%	100,0%

La interpretación de la *tabla* en un informe de resultados requiere que se simplifique, a modo de la *tabla* que figura a continuación. Ésta sólo incluye los porcentajes calculados en el sentido de la variable *edad* (que actúa como variable *independiente*), con las *bases muestrales* correspondientes. El redondeo porcentual hace que en dos agrupaciones (30-49 y 65 y más) sumen 101 y no 100. En la *tabla* puede observarse la existencia de una pauta continua. La percepción en *exceso* de la presencia de inmigrantes aumenta con la edad, con una diferencia porcentual en el gradiente de veinte puntos entre los españoles menores de 30 años (39%) y las personas de 65 y más (59%). Mientras que la valoración de su número de “elevado” e incluso “aceptable” varía en orden inverso: aumenta conforme desciende la *edad* del encuestado. De lo que se deduce la existencia de covariación entre ambas variables: la *percepción de la presencia de inmigrantes* varía con la *edad* y en el sentido esperado de una mayor percepción entre las personas de más edad. Existe un gradiente regular en las respuestas a ambas preguntas, con máximos y mínimos acusados. La graduación exacta de dicha relación de dependencia entre ambas variables precisa del cálculo de *coeficientes de contingencia* y sus correspondientes pruebas de significatividad estadística.

CUADRO 8.4. *Percepción de la presencia de inmigrantes en España de acuerdo con la edad del encuestado*

Encuesta CIS-OBERAXE 2010 (% verticales) Número de inmigrantes	Edad				
	Menos de 30	30-49	50-64	65 y más	Total
Insuficiente (se necesitan más)	-	1	-	1	-
Aceptable	21	22	15	10	17
Elevado	40	35	33	31	35
Excesivo	39	43	52	59	48
Base muestral	494	996	607	598	2.695

La *tabla* se complementa con estadísticos que cuantifican el grado de relación entre las variables y su significatividad. De manera especial, cuando se quiere deducir una relación causal entre las variables comprendidas en la tabla. Cada vez son más los estadísticos incorporados en los paquetes al uso. El programa SPSS, con el que se han elaborado las tablas anteriores, incluye estadísticos de:

- *Bondad de ajuste*: las pruebas de *chi-cuadrado* de Pearson, de razón de *verosimilitud* (que se basa en la teoría de máxima verosimilitud y llega a conclusiones similares a la anterior en tamaños muestrales elevados), de *Mantel-Haenszel* (de asociación lineal por lineal; a partir de variables ordinales) y la *prueba de McNemar* (para detectar cambios en las respuestas debidos a la intervención experimental en los diseños con dos mediciones: antes y después del tratamiento). Permiten comprobar la significatividad de la relación entre las variables; sus posibilidades de inferencia estadística al conjunto de la población (cuando la significación es menor o igual al valor de referencia habitual de 0,05).
- *Medidas direccionales* de la relación de causalidad (grado al que la variable *independiente* explica la variabilidad de la variable *dependiente*). Las más comunes son *lambda* y *tau* de Goodman y Kruskal, junto con el *coeficiente de incertidumbre*, porque pueden emplearse con cualquier tipo de variable. Le siguen los estadísticos *d* de Somer, que precisa que las variables sean al menos *ordinales*, y *eta*, para variables dependientes de *intervalo* [las diferentes tipologías de variables y sus niveles de medición se describen en el capítulo 2].
- *Medidas simétricas* para conocer el grado de relación entre las variables, sin diferenciar entre dependientes e independientes. *Phi*, *V* de Cramer y el *coeficiente de contingencia*, pueden aplicarse con cualquier tipo de variable. El nivel de medición *ordinal* se precisa para la utilización de los estadísticos *tau-b*

y *tau-c* de Kendall, *gamma* de Goodman y Kruskal y el coeficiente de correlación de Spearman. Y para el coeficiente de correlación *r* de Pearson, ambas variables han de estar medidas a nivel de *intervalo*.

Éstos son estadísticos de uso común, aunque no son los únicos, como ya muestra el libro clásico de Siegel (1985). En él se comentan éstos y otros estadísticos cuya explicación sobrepasa los propósitos de la presente publicación. Para su conocimiento se remite, al lector interesado, a la consulta de bibliografía estadística especializada. Sólo añadir que, en las versiones de SPSS para *Windows*, se informa del valor de las medidas direccionales y simétricas y de su significatividad (mediante la prueba *t* de Student). El valor de los coeficientes de contingencia suele oscilar entre “0,0” (que denota inexistencia de relación entre las variables) y “1,0” (asociación perfecta), con signo positivo o negativo. El signo *positivo* expresa que la asociación es “positiva”: ambas variables avanzan en la misma dirección (los aumentos en el valor de una variable suponen incrementos en el valor de la otra). El signo *negativo* denota, en cambio, que ambas variables avanzan en direcciones contrarias (al aumentar el valor de una, descende el valor correspondiente de la otra). A partir de “0,5” (con signo positivo o negativo) la relación se estima relevante; en especial cuando sobrepasa el valor de “0,7” y se aproxima a “1,0”. Si bien, como todo estadístico muestral, su interpretación dependerá de su significatividad. Ésta se comprueba mediante la *razón t*, que se obtiene dividiendo el valor de la medida direccional o simétrica de contingencia entre su *error típico de estimación*. Para que sea significativa, la probabilidad de error en la inferencia ha de ser pequeña. El nivel usual de significatividad es inferior o igual a “0,05”. Expresa una probabilidad de error en la inferencia de la estimación muestral al correspondiente parámetro poblacional del 5%.

Hay que precisar que ni las diferencias porcentuales, ni los estadísticos de contingencia correspondientes, son suficientes para afirmar que existe relación de dependencia entre dos variables. Ello requiere haber anulado la existencia de explicaciones alternativas a la relación causa-efecto analizada, controlando los efectos de variables que puedan mediar en dicha relación.

Por ejemplo, en la relación antes analizada entre *edad* y *percepción del número de inmigrantes*, controlar por la variable *género*; ver si hay diferencias en las respuestas de varones y mujeres. Las diferencias porcentuales en cada categoría de la variable *dependiente* para cada valor de la *independiente* han de mantenerse en cada categoría diferenciada en la variable de *control*. De no ser así, la relación será espúrea y la tercera variable en realidad actuaría como predictora en la relación causa-efecto que se analiza.

La tabla trivariada (cuando se introduce una variable de *control*) tiene una tabla bivariada de la variable *dependiente* e *independiente* para cada categoría de la variable de *control* (una para varones y otra para mujeres). Estas tablas se llaman “parciales”. Su número dependerá del número de categorías que se distinguen en la

variable de *control* (el menor posible). En el ejemplo anterior habría dos subtablas, pero si el número de categorías aumenta a cuatro o más, la interpretación se complica. También cuando además se controla por una “cuarta” variable. En tales casos la mejor opción será elegir alguna de las técnicas analíticas multivariantes que permiten analizar conjuntamente los efectos de dependencia entre más de dos variables.

8.4. Análisis multivariable

El análisis conjunto de las interrelaciones y covariaciones entre más de dos variables a través de técnicas analíticas multivariantes muestra mayor adecuación para el análisis de la compleja realidad social. Los notorios avances en la accesibilidad de los programas estadísticos está contribuyendo a que los informes de investigación no se limiten a meros comentarios porcentuales univariantes o bivariantes, sino que prosigan con modelos estadísticos multivariantes que permitan medir, con una mayor precisión, los efectos de cada variable.

Existe un amplio abanico de técnicas analíticas multivariantes. Su elevado número, unido a la complejidad de los análisis, hacen inviable que este apartado ofrezca una exposición detallada de cada técnica. Se ofrece apenas una visión panorámica genérica de las características que distinguen a las principales técnicas analíticas. Al lector interesado se le remite a que consulte manuales especializados en análisis multivariantes. Algunos están referenciados en el apartado correspondiente a lecturas complementarias y en la bibliografía.

Existen diferentes clasificaciones del amplio abanico de técnicas analíticas multivariantes. La más compartida es aquella que distingue las técnicas de *dependencia* y de *interdependencia*. Las primeras analizan relaciones de causalidad (o dependencia). Las segundas indagan en relaciones de *interdependencia*, en la posibilidad de agrupación de variables o casos en función de sus covarianzas o correlaciones, sin diferenciar entre variables dependientes e independientes. Para facilitar la visualización conjunta de las técnicas, la figura 8.1 resume las técnicas analíticas de *dependencia* y la figura 8.2 las de *interdependencia* (y que después se detallan).

A) Regresión lineal múltiple

Analiza la relación de dependencia entre dos o más variables independientes o (predictoras) y una dependiente, que ha de ser *métrica* (o continua). El análisis persigue un triple objetivo:

- a) Predecir el valor promedio de la variable *dependiente* (Y) a partir de valores concretos de las *independientes* (X) mediante una *ecuación lineal*.
- b) Cuantificar el grado de relación entre las variables, mediante el coeficiente de *correlación R de Pearson* y su cuadrado, el coeficiente de *determinación*.

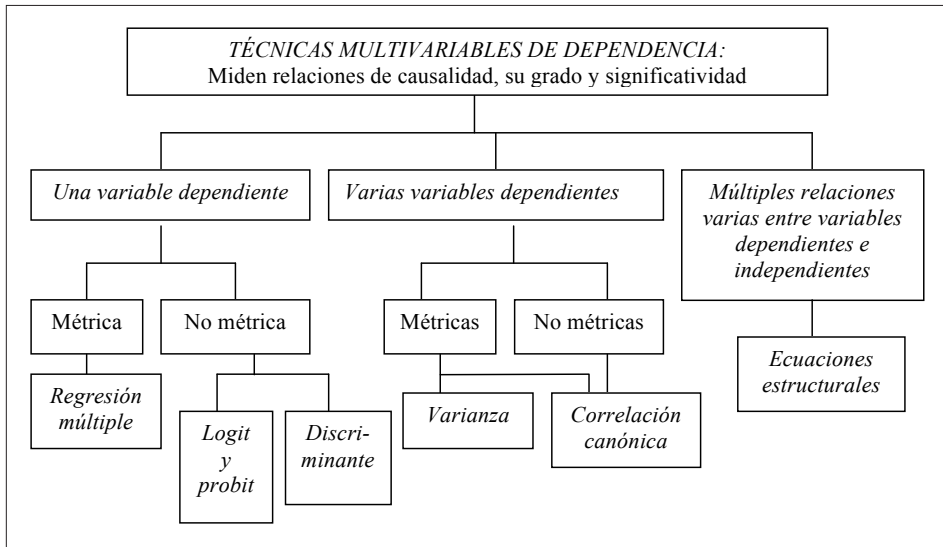


FIGURA 8.1. Técnicas de análisis multivariable de dependencia.

Éste expresa en qué proporción el valor de Y es determinado por las X en la ecuación.

- c) Determinar el grado de confianza al que puede afirmarse que la relación observada en los datos muestrales es realmente cierta (la significatividad del modelo de regresión en su conjunto a través del estadístico F , y de cada coeficiente de regresión en particular mediante la t de *Student*).

Cuando el número de variables independientes se reduce a una, el análisis es de *regresión lineal simple*. Su correcta realización exige el cumplimiento de una serie de *supuestos* básicos:

- a) *Tamaño muestral* elevado (preferiblemente 40 casos por cada variable independiente).
- b) Variable dependiente *métrica-continua*.
- c) Variables independientes *métricas o continuas* y relevantes para la predicción de la variable dependiente.
- d) *Linealidad*: la relación entre la variable dependiente y las independientes ha de ser “lineal” (el efecto de cada independiente en la dependiente es el mismo cualquiera que sea el valor de la independiente).
- e) *Aditividad*: los efectos de las variables independientes en la dependiente han de poderse sumar entre sí.
- f) *Normalidad*: correspondencia de la distribución de datos con la curva normal para favorecer la realización de las pruebas de significatividad.

- g) *Homocedasticidad*: la varianza de la variable dependiente ha de ser igual en cada valor de la independiente para que la relación de las independientes con la dependiente pueda medirse con rigor. En caso contrario, habría una mejor predicción del valor de la dependiente en unos determinados valores de las independientes que en otros.
- h) *Ausencia de colinealidad* o de correlación entre las independientes para que puedan medirse los efectos concretos de cada independiente en la dependiente.
- i) *Independencia de los términos de error*: el valor de la variable dependiente en cada caso concreto ha de ser independiente del resto.

El incumplimiento de estos supuestos lleva a la aplicación de procedimientos de regresión alternativos, como la *regresión de mínimos cuadrados ponderados* (cuando se incumplen los supuestos de *homocedasticidad* e independencia de los términos de error) y la *regresión polinómica* (si se incumple el supuesto de *linealidad*). En este último caso, también se aplicaría la *regresión logística*.

B) *Análisis logit*

Como el análisis *probit*, constituyen modelos analíticos alternativos para la predicción de variables dependientes *cualitativas* (o no métricas). Si bien, el análisis *logit* ha alcanzado un desarrollo más amplio por su mayor practicabilidad. Incluye la *regresión logística* y el *modelo logit*.

- *Regresión logística*: una alternativa idónea cuando la variable dependiente es categórica o no métrica. Los *objetivos* son comunes a los análisis de regresión, aunque difiere su configuración y estadísticos:
 1. Predecir la *probabilidad* de ocurrencia de un determinado evento ($Y = 1$), a partir de los valores que presentan una serie de variables independientes *continuas*.
 2. Cuantificar el grado de relación entre las independientes y la dependiente (mediante los coeficientes R^2 de Nagelkerke o de razón de verosimilitud (entre otros)).
 3. Determinar la significatividad del modelo de predicción (χ^2 de razón de verosimilitud) y de los coeficientes por separado (W de Wald o t de Student), entre otros.

No exige el cumplimiento de los supuestos de *linealidad*, *normalidad* y *homocedasticidad*. Sí los otros supuestos antes referidos: tamaño muestral elevado (mínimo 15 casos por variable predictora), variables independientes relevantes y continuas, ausencia de *colinealidad* y *aditividad*.

- *Modelo logit*: deriva del modelado *log-linear*, pero aplicado al estudio de la causalidad. Analiza la relación causa-efecto entre una única variable dependiente y varias independientes o predictoras. Su denominación “logit” proviene del cálculo del *logaritmo de la razón de probabilidades* de los casos donde acontece un determinado evento respecto de aquellos donde no se presenta. A diferencia de la *regresión logística*, la predicción de la probabilidad de ocurrencia de la variable dependiente puede hacerse con variables predictoras continuas y/o categóricas.

C) Análisis discriminante

Analiza la relación entre una única variable dependiente (no métrica o categórica) con varias independientes (métricas). La variable dependiente la forman los grupos de casos o variables previamente configurados, por alguna técnica exploratoria de clasificación (como el análisis de *conglomerados* o *factorial*), aunque no siempre. Los *objetivos* principales son:

- a) Comprobar si una determinada agrupación de objetos es correcta y si queda adecuadamente caracterizada por las variables que definen cada grupo.
- b) Averiguar cuál es la combinación lineal de variables predictoras (las *funciones discriminantes*) que hace máxima la diferencia entre los grupos.
- c) Predecir la probabilidad de pertenencia de un objeto concreto (caso, variable) a uno de los grupos diferenciados, a partir de los valores que presente en las variables predictoras. Como en el análisis *logit*, se emplea la razón de aciertos de la *tabla de clasificación* como una medida del éxito del modelo en la predicción de la variable dependiente (los grupos diferenciados en función de la variable de interés. *Por ejemplo*, su rendimiento académico o su actitud ante la inmigración).

Comparte los mismos supuestos de *regresión lineal múltiple*: tamaño muestral elevado (mínimo 20 casos por variable independiente), variables independientes continuas, ausencia de *colinealidad*, *normalidad* y *homocedasticidad* o igualdad de las matrices de covarianza en cada grupo.

D) Análisis multivariable de la varianza (MANOVA)

Una técnica analítica de amplia aplicación en la investigación *experimental* como ha podido observarse en el capítulo 3. A ello contribuye la especificidad de su análisis. Explora de manera simultánea las relaciones entre varias variables independientes no métricas (normalmente referidas como *tratamientos*) y dos o más dependientes métricas. El objetivo es determinar la existencia de diferencias en los va-

lores medios de las variables dependientes, en cada grupo de *tratamiento*. Ello exige el cumplimiento de tres *supuestos* básicos:

1. Las variables dependientes han de estar distribuidas normalmente (*normalidad*).
2. Matrices de *varianzas-covarianzas* iguales en todos los grupos (*homocedasticidad*). En caso contrario, no podría medirse el efecto específico del *tratamiento*, porque se partiría de grupos inicialmente diferentes.
3. Tamaño muestral elevado. Ha de superar al necesario para la varianza simple, si se pretende que el modelo adquiera significatividad estadística.

El análisis de *varianza simple* (ANOVA) comprueba la existencia de diferencias grupales respecto a una única variable *dependiente* (*métrica*). Si se observan diferencias entre las *medias grupales*, después se procede a la comparación de las *varianzas grupales* y a la medición de su *significatividad*. Interesa que la *varianza* entre los grupos supere a la *varianza intragrupal*. Ello expresaría una mayor heterogeneidad entre los grupos, frente a una escasa variabilidad dentro de ellos. Por lo que podría afirmarse que existen diferencias estadísticamente significativas entre los grupos. La *significatividad* de las diferencias grupales se comprueba mediante los estadísticos “*t*” (si únicamente se han formado dos grupos de tratamiento) y “*F*” (si son más de dos los grupos creados). En ésta, como en cualquier prueba de *significatividad*, se comparan los valores *empíricos* (“*t*” y “*F*”) con los *teóricos* (mostrados en las tablas de la “*t*” de Student y de la “*F*” de Fisher, correspondientes), al igual que en toda prueba de significatividad.

E) Correlación canónica

Una técnica de *dependencia* que permite comprobar la existencia de interrelaciones entre una serie de variables *dependientes* y otra serie de variables *independientes* o predictoras. Las variables pueden ser métricas o no métricas. La finalidad principal es obtener las *combinaciones lineales* de cada serie de variables (dependientes e independientes) que maximicen las correlaciones entre ellas. El grado de relación entre la serie de *variables canónicas* se mide mediante el coeficiente de *correlación canónica*. Éste, elevado al cuadrado (R^2), representa el porcentaje de *varianza* de una combinación de variables *dependientes canónicas* que es explicada por una combinación de las variables *independientes*. Su *significatividad* se comprueba, de nuevo, mediante el estadístico “*F*”.

Las contribuciones de cada variable (*dependiente* e *independiente*) a la combinación o serie de *variables canónicas* respectiva se comprueba, al igual que en el *análisis factorial* y *discriminante*, mediante los *factores de carga* (“*canonical loadings*”). Éstos han de ser “ $\geq 0,30$ ” para que se consideren relevantes. La contribución de las variables también puede comprobarse mediante los *pesos canónicos*

(“*canonical weights*”), si bien estos últimos presentan una mayor inestabilidad que los anteriores (los *factores de carga*).

La *correlación canónica* comparte supuestos y estadísticos presentes en los análisis de *regresión lineal múltiple* y *discriminante*.

F) Ecuaciones estructurales

Término genérico que designa diferentes aproximaciones al análisis de la causalidad entre una serie de variables independientes y dependientes, con la particularidad de incluir distintas interrelaciones entre ellas y diferenciar entre variables *observadas* y *latentes*. Proporciona un modelo causal más ajustado a la pluridimensionalidad de las relaciones causales, que combina los análisis de *regresión lineal múltiple*, *factorial* y de *sendero* (que contribuye a la representación gráfica de las relaciones causales). Éstas se representan gráficamente mediante un diagrama de *sendero* (“*path*”), distinguiendo entre variables *latentes*, *observadas* y *errores de medición* (de cada modalidad de variable), y el tipo de relación entre ellas (*por ejemplo*, una flecha bidireccional indica existencia de correlación entre dos variables). Para cada variable que recibe una flecha se formula una ecuación. En esa ecuación, dicha variable actúa como variable *dependiente*. Las diferentes relaciones causales tendrán sus correspondientes “ecuaciones lineales estructurales”.

Los *parámetros* se obtienen mediante el método de *mínimos cuadrados generalizados*, o por el de *máxima verosimilitud*. Si bien, el primer método muestra imprecisión, conforme desciende el tamaño de la *muestra* y aumenta la complejidad del modelo. La *significatividad* de los *parámetros* se comprueba mediante la “*t*” de *Student*, además de considerarse el *error* de la estimación.

Como se diferencia entre variables *latentes* y *observadas*, se comprueba la *fiabilidad* ($\geq 0,70$) y la *validez* ($\geq 0,50$) de la *medición* de los *constructos teóricos* (*variables latentes*). A ello se suma el *ajuste* global del modelo alcanzado respecto al inicial. Entre los *índices de ajuste* destacan el estadístico χ^2 (aunque ahora interesan los valores bajos, porque expresan un mejor *ajuste* entre el modelo obtenido y el propuesto inicialmente); el *índice GFI*, que oscila entre “0,00” y “1,00” (cuanto más se aproxime a “1,00”, mejor); los *índices TLI* ($\geq 0,90$), *NFI* ($\geq 0,90$), *AGFI* ($\geq 0,90$), y *AIC* ($\geq 0,70$), entre otros.

El análisis de *ecuaciones estructurales* comparte los *supuestos* básicos señalados en el análisis de *regresión lineal múltiple*. A decir: *normalidad*, *linealidad* de las relaciones entre variables, *aditividad* de los efectos, observaciones independientes, variables dependientes e independientes continuas, ausencia de *colinealidad* entre las independientes y *muestra* elevada aleatoria.

A las técnicas de *dependencia* se suman las de *interdependencia*. Éstas analizan relaciones simétricas entre variables (sin diferenciar entre dependientes e independientes) o casos. Les une un mismo *objetivo* básico: la agrupación de casos o variables con un nexo común (que compartan variabilidad). No obstante existen diferencias sustan-

tivas entre las diferentes técnicas. La principal (no la única) atiende al nivel de medición que han de cumplir las variables para la correcta realización de los análisis. La figura 8.2 reúne a las principales técnicas analíticas multivariantes de *interdependencia*.

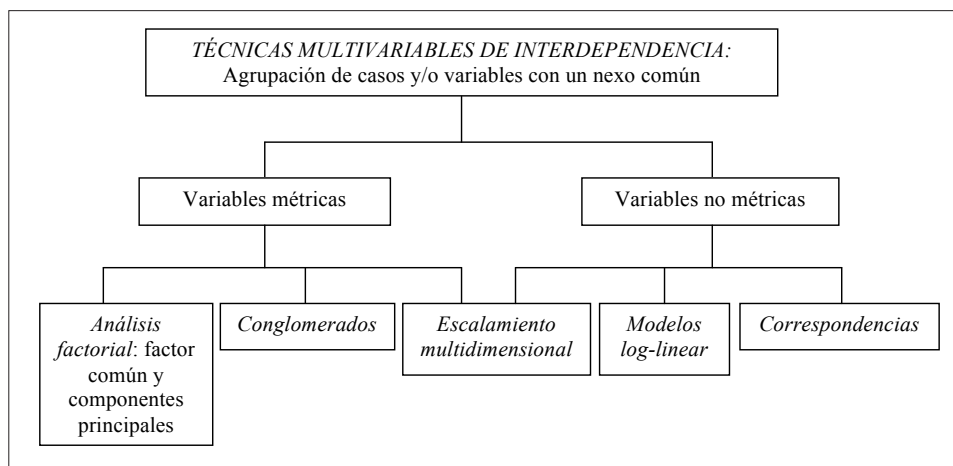


FIGURA 8.2. Técnicas de análisis multivariable de interdependencia.

A) *Análisis factorial*

Aglutina diferentes procedimientos analíticos que indagan en la varianza compartida por una serie de variables en busca de la estructura “latente” (no directamente observable) que las vincula. Entre sus *objetivos* principales están:

- a) Sintetizar la información que proporciona una serie de variables observadas (*indicadores*) en un número inferior de variables *latente* (no observadas), llámense *factores comunes* o *componentes principales*, en función del tipo de análisis que se aplique. Esta serie de variables *latentes* han de aglutinar *indicadores* bastante correlacionados entre sí y poco correlacionados con *indicadores* que forman otra estructura latente. La “no correlación” entre variables latentes es una propiedad importante. Significa que los *indicadores* miden “dimensiones” diferentes en los datos.
- b) La obtención de *puntuaciones factoriales*, que actúen en representación de los *factores* o *componentes* en análisis posteriores (como *discriminante*, *regresión* o *varianza*).

El *análisis factorial* puede ser exploratorio o confirmatorio. El *exploratorio* busca la configuración inicial de los factores o componentes, mientras que el *confirmatorio* la corroboración de un modelo inicial de agrupación de indicadores en dimensiones latentes. Dentro de los análisis *exploratorios* se diferencia el análisis de

componentes principales (que analiza la varianza total del conjunto de variables observadas) y los de *factor común* (se restringen a la varianza “común” o compartida por la serie de indicadores que analizan). En este último grupo están los análisis de: *factor principal* o de ejes principales, de *máxima verosimilitud*, de *mínimos cuadrados generalizados* y *no ponderados*, *factorización alfa* y *factorización imagen*.

También se distingue entre el análisis *factorial R* (de variables) y el *Q* (de casos). En los últimos años han aparecido además programas estadísticos, como PRINCALS, que permiten la consecución de modelos factoriales *no métricos*, desbancando la concepción tradicional “métrica” del *análisis factorial*. Ésta exige el cumplimiento de cuatro *supuestos* básicos: tamaño muestral elevado, *normalidad* multivariable, *linealidad* y *correlaciones* elevadas entre las variables (para que pueda procederse a su agrupación de acuerdo con la varianza que comparten).

B) Análisis de conglomerados

Igualmente persigue la agrupación de casos o variables en un número reducido de grupos, ahora llamados “conglomerados”. La condición es que éstos sean mutuamente excluyentes: que los casos o variables que formen un conglomerado sean lo más similares entre sí (respecto al criterio de clasificación que se aplique) y diferentes de los integrantes de los otros *conglomerados*.

Existe una amplia diversidad de análisis de *conglomerados*. Se clasifican en métodos de conglomeración *jerárquicos* y *no jerárquicos*. Los primeros se dirigen a la búsqueda de *conglomerados* mediante procedimientos *aglomerativos* (el análisis comienza con tantos conglomerados como objetos a clasificar; concluye con un solo conglomerado, que agrupa a todos ellos) o *divisivos* o de participación (se parte de un único conglomerado, que se va dividiendo hasta obtener tantos conglomerados como objetos a clasificar). Por el contrario los métodos de conglomeración *no jerárquicos* (o de optimización) buscan “optimizar” el criterio de selección inicial: parten de una clasificación previa. Dentro de estos últimos, el más aplicado es el análisis *K-medias*. El análisis parte de una partición inicial de los datos. Los casos o variables se clasifican en el conglomerado a cuyo centro (o *centroide*) se hallen más próximos, a través de un procedimiento iterativo (de reestimación continua) hasta alcanzar la mejor clasificación posible: conglomerados lo más divergentes posible entre sí. Esta variedad analítica (*K-medias*) se adecua a tamaños muestrales grandes, mientras que los *no jerárquicos* a muestras pequeñas. Ésta es una de las razones principales de su mayor popularidad entre los análisis de *conglomerados*.

C) Escalamiento multidimensional

Al igual que el análisis *factorial*, su objetivo principal es la obtención de un número reducido de *dimensiones*, que permitan caracterizar a los objetos a ser clasifi-

cados. Pero difiere en que impone la restricción sobre el número de *dimensiones* a extraer. Han de ser pocas para que las agrupaciones que resulten puedan representarse gráficamente. Los datos se representan como puntos en un espacio “multidimensional”. La distancia entre ellos se considera una expresión gráfica de su semejanza o disimilaridad. La representación gráfica es más factible e interpretable cuando la información se reduce a dos o tres dimensiones.

En función del nivel de medición de las variables, se diferencia entre escalamiento multidimensional “métrico” o “no métrico”. En el primer caso la clasificación se hace a partir de la *matriz de correlaciones* entre las variables; en el segundo, desde una *matriz de rangos*. De ella se obtiene información de la *similaridad* de los objetos. La configuración final será aquella que represente mejor a los *rangos* de la matriz inicial. Como el análisis *factorial*, el *escalamiento multidimensional* puede tener carácter *exploratorio* o *confirmatorio*, de acuerdo con la finalidad que persiga el análisis.

D) Modelos “log-linear”

También llamados modelos “lineales logarítmicos” o “log-lineales”. El nombre proviene de la transformación logarítmica de los datos (las frecuencias observadas de las variables) para facilitar su *aditividad*: la posibilidad de sumar los distintos efectos de un conjunto de variables *no métricas* (a semejanza de las variables métricas en el análisis de *regresión*). Entre sus *objetivos* principales están:

- a) Analizar las relaciones entre variables (cualitativas), representadas en *tablas de contingencia multidimensionales*.
- b) Estimar la probabilidad de obtener un determinado valor en una celdilla concreta de la *tabla*.

Las *tablas de contingencia* muestran gran utilidad para el análisis de variables *cualitativas* o *no métricas*. Pero, la incorporación de una tercera, cuarta e incluso quinta variable (como variables de *control*) complica su interpretación. Para ello se recomienda el uso preferente de los modelos *log-linear* y *logit*, dependiendo de la finalidad del estudio: el *análisis logit* para relaciones de *dependencia*, mientras que el *modelado log-linear* para las de *interdependencia*. En ambos análisis, las celdillas de la tabla se traducen a componentes, llamados parámetros *lambda*, que miden el efecto de los valores que forman cada celdilla. Éstos serán relevantes cuando su significatividad sea menor o igual a 0,05 (corresponde a un valor $Z \geq 1,96$), como es usual. Permiten conocer la probabilidad de que cada caso comparta una combinación específica de atributos de un conjunto de variables.

Como en las *tablas de contingencia*, el ajuste del modelo se realiza a través de las pruebas de *chi-cuadrado* y de *razón de verosimilitud*, cuyos valores coinciden a medida que aumenta el tamaño de la *muestra*.

E) Análisis de correspondencias

Una técnica analítica que también se adecua al análisis de variables cualitativas. Parte, igualmente, de la configuración de las variables en una *tabla de contingencia*. Pero difiere en su objetivo principal: la representación de las distancias de las *filas* y de las *columnas*, que componen la tabla, en unos ejes cartesianos. Para ello se transforma el valor de *chi-cuadrado* en una medida métrica de distancia.

Comparte con el *escalamiento multidimensional* que la proximidad de los puntos mide la *similaridad* o asociación entre las categorías de las variables. Como en otras técnicas de *interdependencia*, se identifica el número de dimensiones que categorizan al objeto de estudio (mediante el porcentaje acumulado de varianza explicada). Para cada dimensión se derivan los *autovalores* (que miden la contribución relativa de cada dimensión en la explicación de la varianza de las variables) y se identifica la asociación o relación entre las categorías de las variables mediante su “proximidad”. Ésta se comprueba en las *filas* o las *columnas*, dependiendo de su ubicación. El obtener un valor *chi-cuadrado* elevado significa que existe un grado fuerte de “correspondencia” entre los atributos de las variables. Lo cual permite la reducción dimensional de las proporciones de objetos en una serie de atributos. Al mismo tiempo, se consigue la representación de los objetos relacionados con esos atributos.

8.5. Informe de la investigación: contenidos básicos y redacción

La investigación concluye con la redacción del *informe*. Éste cumple la función de comunicar a la población en general, y a la comunidad científica en particular, los hallazgos de la investigación desde su inicio a su término; de modo que cualquier investigador pueda replicar el estudio. Como señalara Caplow (1977: 15), “El objetivo oficial del investigador es siempre el de contribuir al enriquecimiento del conocimiento científico. Se sobrentiende que aquél publicará los resultados obtenidos sin restricciones ni deformaciones, y que proporcionará suficientes informaciones sobre sus métodos a fin de que otros investigadores puedan reproducir este estudio si así lo desean”. Y ello de la forma más objetiva, transparente y fidedigna posible de los contenidos de la investigación.

“El informe tiene que dar cuenta del cómo, cuándo y con qué resultados se han cumplido los objetivos del proyecto. Un informe debe explicar los procesos seguidos, las divergencias o cambios en lo proyectado (si los hubiere) y la confirmación o refutación de las hipótesis de partida que dieron lugar al proyecto de investigación” (Alvira, 1995: 246-247).

La estructura del *informe* se ajusta a las diferentes fases de la investigación. Puede adoptar una variedad de formas, dependiendo del público al que se dirige

(profano, experto), el medio empleado para su divulgación (informe técnico, monografía, artículo de revista científica, ponencia o comunicación de congreso...), y el ingenio y estilo comunicativo del investigador. Si bien suele mantener una estructura básica común, cuyo desarrollo varía en función de sus características. Seis son los apartados principales que lo componen:

- a) *Introducción*: el informe suele comenzar con una introducción que ofrece una visión global de la investigación y su relevancia para el conocimiento del tema de estudio. El investigador define y contextualiza el *problema* de investigación, especificando los *objetivos* e *hipótesis* esenciales, a partir de indagaciones teórico-empíricas precedentes. Se resalta el interés y trascendencia del estudio para el avance del conocimiento sobre lo que se analiza, conectándolo con la investigación que le antecede. El propósito no es mostrar un catálogo de la literatura especializada sobre el tema, sino señalar cuál es el estado actual de la cuestión: qué cuestiones se han resuelto y sobre cuáles existen aún controversias o carencia de estudios. Dicha argumentación se acompaña de un número limitado de referencias bibliográficas seleccionadas para marcar la trayectoria investigadora que la precede.
- b) *Metodología*: descripción del diseño y desarrollo de la investigación. Su grado de detalle y ubicación dependerá del tipo de informe. En los artículos científicos el detalle metodológico suele ubicarse después de la *introducción*, mientras que en los *informes* de carácter más divulgativo (enfocados a la exposición de resultados) puede ir al final como “apéndice metodológico”.

El apartado correspondiente a la *metodología* especifica la *estrategia* de investigación que se ha seguido para alcanzar los *objetivos* de la investigación y comprobar las *hipótesis* principales que guían el estudio. Se detallan las *técnicas* aplicadas para la obtención y el análisis de la información, además de especificidades relativas a la *muestra* diseñada y la que resulta del *trabajo de campo*. A ello se añade cualquier información que contribuya a la interpretación de los resultados de la investigación: problemas encontrados en la materialización del *diseño* proyectado, actuaciones para su resolución y límites posibles del estudio (su *validez* y *fiabilidad*).

- c) *Resultados*: exposición e interpretación de los “hallazgos” de la investigación, de un modo claro, preciso y objetivo. El texto suele acompañarse de tablas y gráficos que sintetizan los datos más relevantes y ayuden a su exposición.
- d) *Discusión*: los *resultados* se interpretan relacionándolos con las *hipótesis* de la investigación. Se discute si son o no corroboradas por la información reunida y se comparan con los de estudios precedentes (si los hubiere), precisando la relevancia y límites a su generalización. De esta reflexión se deducirán respuestas a algunas cuestiones planteadas en la *Introducción* del *Informe*. Otras quedarán sin respuesta, poniendo de manifiesto los límites del estudio. Por lo que es común que este apartado concluya con la propuesta de futuras líneas de investigación.

- e) *Conclusión*: este apartado es, como señala Gilbert (1993: 340), “el más difícil de escribir porque debería resumir cuáles fueron los principales hallazgos de la investigación de una forma sucinta e interesante. Esto es, porque muchos lectores mirarán primero el final para ver si tiene algo interesante que decir, antes de comenzar por el principio”. En este apartado se resaltan los hallazgos más significativos del estudio, dejando la puerta abierta a futuras indagaciones que resuelvan cuestiones discutidas o que han quedado sin respuesta en la investigación llevada a cabo.
- f) *Bibliografía*: la relación completa de las referencias bibliográficas de los estudios citados a lo largo del informe, ordenados alfabéticamente y por año de publicación. Las normas para citar las referencias dependen de la revista, editorial o colección donde se publique el informe. En general las referencias de *libros* incluyen: el apellido principal del autor y las iniciales del nombre, la fecha de publicación de la obra (entre paréntesis), el título del libro (en cursiva), la editorial y la ciudad donde se publica. Para las *revistas*, el título del artículo aparece entre comillas, acompañado del nombre de la revista (en cursiva), el número o volumen de la misma y las páginas donde aparece en la revista. *Por ejemplo*, véase cómo aparecen citadas las referencias bibliográficas a lo largo del presente libro.

A este esquema genérico suele añadirse un breve *resumen* con los hallazgos más relevantes del estudio (normalmente al inicio del *Informe*, aunque también puede ir al final). Cumple la función de ofrecer una visión sintética global de la investigación (entre 100 y 250 palabras), que despierte el interés del lector. Y para facilitar su localización en bases de datos, suele acompañarse de una selección de palabras claves (cuando adquiere la forma de artículo de revista científica). E indistintamente de su formato, el apartado de *agradecimientos* o mención de las personas e instituciones que contribuyeron en la realización de la investigación es habitual (entidades que financiaron o colaboraron en su materialización, y particulares que ayudaron al equipo investigador o que aportaron sugerencias a su diseño, ejecución y redacción).

Lecturas complementarias

- Alexander, C. (1986): *Writing for social scientists: how to start and finish your thesis, books or article*. Chicago University Press. Chicago.
- Biemer, P. P. y Sharon, L. Ch. (2008): “Weighting survey data”, en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International Handbook of Survey Methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 317-341.
- Bryman, A. y Cramer, D. (1995): *Quantitative data analysis for social scientist*. Routledge. Londres.
- Cea D’Ancona, M.^a Á. (2002/2008): *Análisis multivariable. Teoría y práctica en la investigación social*. Síntesis. Madrid.
- Díaz de Rada, V. (2009): *Análisis de datos de encuestas*. Editorial UOC. Barcelona.
- Eco, H. (1982): *Cómo se hace una tesis*. Gedisa. Barcelona.

- Gilbert, N. (1993): "Writing about social research", en Gilbert, N., *Research social life*. Sage. Londres: 329-344.
- De Miguel, A. (1997): *Manual del perfecto sociólogo*. Espasa. Madrid.
- Hair, J. F.; Anderson, R. E.; Tathan, R. L.; Black, W. C. (1999): *Análisis multivariante*. Prentice Hall. Madrid.
- Lévy Manguin, J. P. y Varela Mallou (comps.) (2003): *Análisis multivariable para las ciencias sociales*. Pearson Education. Madrid.
- Pardo Merino, A. y Ruiz Díaz, M. A. (2005): *Análisis de Datos con SPSS 13 Base*. McGraw-Hill. Madrid.
- Ritchey, F. J. (2008): *Estadística para las Ciencias Sociales*. McGraw-Hill. México.
- Sánchez Carrión, J. J. (1995/1999): *Manual de análisis de datos*. Alianza. Madrid.

Ejercicios propuestos

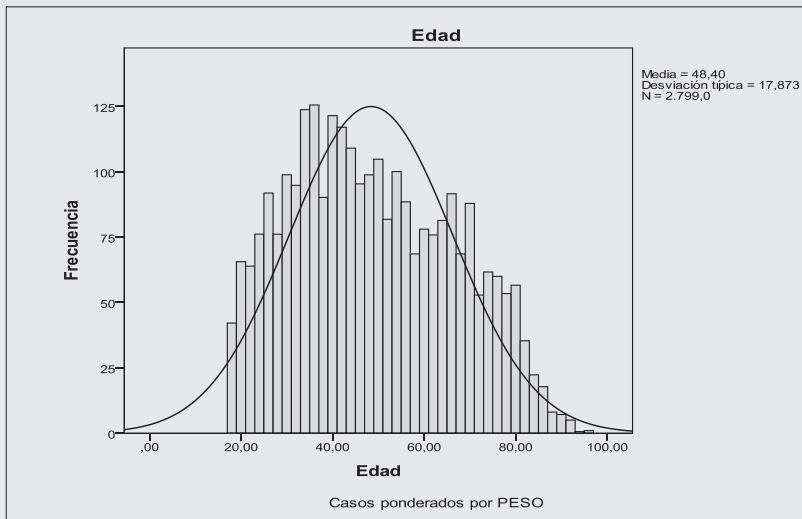
- Interprete los datos siguientes sabiendo que la variable *ideología política* va en una escala de 0 (extrema izquierda) a 10 (extrema derecha); *edad*, en años cumplidos; *conveniencia de personas de diferentes culturas* en una escala de 0 (muy negativo) a 10 (muy positivo); y *número de inmigrantes* de 4 (excesivo), 3 (elevado), 2 (aceptable) y 1 (insuficiente). Los datos corresponden a la encuesta CIS-OBERAXE de septiembre de 2010 (Cea D'Ancona y Valles, 2011).

Número de inmigrantes

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00	11	,4	,4	,4
	2,00	466	16,6	17,3	17,7
	3,00	936	33,4	34,7	52,4
	4,00	1282	45,8	47,6	100,0
	Total	2694	96,2	100,0	
Perdidos	Sistema	106	3,8		
Total		2800	100,0		

Estadísticos

		Ideología política	Edad	Conveniencia personas diferentes culturas	Número de inmigrantes
N	Válidos	2161	2799	2571	2694
	Perdidos	639	1	229	106
Media		4,5110	48,3970	6,2442	3,2944
Error típ. de la media		,04491	,33782	,04695	,01466
Mediana		5,0000	47,0000	6,0000	3,0000
Moda		5,00	42,00	5,00	4,00
Desv. típ.		2,08760	17,87260	2,38085	,76077
Varianza		4,358	319,430	5,668	,579
Asimetría		,134	,245	-,453	-,607
Error típ. de asimetría		,053	,046	,048	,047
Curtosis		-,001	-,906	,135	-,845
Error típ. de curtosis		,105	,093	,097	,094
Rango		10,00	78,00	10,00	3,00
Percentiles	25	3,0000	34,0000	5,0000	3,0000
	50	5,0000	47,0000	6,0000	3,0000
	75	6,0000	63,0000	8,0000	4,0000



2. A partir de la siguiente tabla de contingencia, elabore una tabla para un informe, interprete los datos y calcule un índice. La variable *número de inmigrantes* está como se definió en el ejercicio anterior. *Ideología política* está recodificada de modo que el valor 1 denota izquierda, 2 centro y 3 derecha.

Tabla de contingencia Número de inmigrantes * Ideología política * P44

				Ideología política			
P44				1,00	2,00	3,00	Total
Hombre	Número de inmigrantes	1,00	Recuento	5	3	0	8
			Frecuencia esperada	2,7	4,1	1,2	8,0
			% N.º inmigrantes	62,5%	37,5%	,0%	100,0%
			% Ideología política	1,4%	,6%	,0%	,8%
		2,00	Recuento	100	90	19	209
			Frecuencia esperada	69,9	106,5	32,6	209,0
			% N.º inmigrantes	47,8%	43,1%	9,1%	100,0%
			% Ideología política	28,1%	16,6%	11,4%	19,6%
		3,00	Recuento	121	199	42	362
			Frecuencia esperada	121,1	184,4	56,5	362,0
			% N.º inmigrantes	33,4%	55,0%	11,6%	100,0%
			% Ideología política	34,0%	36,7%	25,3%	34,0%
		4,00	Recuento	130	250	105	485
			Frecuencia esperada	162,3	247,1	75,7	485,0
			% N.º inmigrantes	26,8%	51,5%	21,6%	100,0%
			% Ideología política	36,5%	46,1%	63,3%	45,6%
	Total		Recuento	356	542	166	1064
			Frecuencia esperada	356,0	542,0	166,0	1064,0
			% N.º inmigrantes	33,5%	50,9%	15,6%	100,0%
			% Ideología política	100,0%	100,0%	100,0%	100,0%
Mujer	Número de inmigrantes	1,00	Recuento	1	2	0	3
			Frecuencia esperada	,9	1,5	,5	3,0
			% N.º inmigrantes	33,3%	66,7%	,0%	100,0%
			% Ideología política	,3%	,4%	,0%	,3%
		2,00	Recuento	90	71	14	175
			Frecuencia esperada	53,1	90,2	31,7	175,0
			% N.º inmigrantes	51,4%	40,6%	8,0%	100,0%
			% Ideología política	28,8%	13,4%	7,5%	17,0%
		3,00	Recuento	107	201	65	373
			Frecuencia esperada	113,2	192,3	67,5	373,0
			% N.º inmigrantes	28,7%	53,9%	17,4%	100,0%
			% Ideología política	34,3%	37,9%	34,9%	36,3%
		4,00	Recuento	114	256	107	477
			Frecuencia esperada	144,8	245,9	86,3	477,0
			% N.º inmigrantes	23,9%	53,7%	22,4%	100,0%
			% Ideología política	36,5%	48,3%	57,5%	46,4%
	Total		Recuento	312	530	186	1028
			Frecuencia esperada	312,0	530,0	186,0	1028,0
			% N.º inmigrantes	30,4%	51,6%	18,1%	100,0%
			% Ideología política	100,0%	100,0%	100,0%	100,0%

			Ideología política				
P44			1,00	2,00	3,00	Total	
Total	Número de inmigrantes	1,00	Recuento	6	5	0	11
			Frecuencia esperada	3,5	5,6	1,9	11,0
			% N.º inmigrantes	54,5%	45,5%	,0%	100,0%
			% Ideología política	,9%	,5%	,0%	,5%
		2,00	Recuento	190	161	33	384
			Frecuencia esperada	122,6	196,8	64,6	384,0
			% N.º inmigrantes	49,5%	41,9%	8,6%	100,0%
			% Ideología política	28,4%	15,0%	9,4%	18,4%
		3,00	Recuento	228	400	107	735
			Frecuencia esperada	234,7	376,6	123,7	735,0
			% N.º inmigrantes	31,0%	54,4%	14,6%	100,0%
			% Ideología política	34,1%	37,3%	30,4%	35,1%
		4,00	Recuento	244	506	212	962
			Frecuencia esperada	307,2	493,0	161,9	962,0
			% N.º inmigrantes	25,4%	52,6%	22,0%	100,0%
			% Ideología política	36,5%	47,2%	60,2%	46,0%
Total	Recuento	668	1072	352	2092		
	Frecuencia esperada	668,0	1072,0	352,0	2092,0		
	% N.º inmigrantes	31,9%	51,2%	16,8%	100,0%		
	% Ideología política	100,0%	100,0%	100,0%	100,0%		

Pruebas de chi-cuadrado

P44		Valor	gl	Sig. asintótica (bilateral)
Hombre	Chi-cuadrado de Pearson	47,396 ^a	6	,000
	Razón de verosimilitudes	47,450	6	,000
	Asociación lineal por lineal	41,557	1	,000
	N.º de casos válidos	1064		
Mujer	Chi-cuadrado de Pearson	52,993 ^b	6	,000
	Razón de verosimilitudes	52,171	6	,000
	Asociación lineal por lineal	42,321	1	,000
	N.º de casos válidos	1028		
Total	Chi-cuadrado de Pearson	95,440 ^c	6	,000
	Razón de verosimilitudes	95,138	6	,000
	Asociación lineal por lineal	84,546	1	,000
	N.º de casos válidos	2092		

a. 3 casillas (25,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 1,25.

b. 3 casillas (25,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es ,54.

c. 2 casillas (16,7%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 1,85.

Medidas simétricas

P44			Valor	Error típ. asint. ^a	T aproximada ^b	Sig aproximada
Hombre	Nominal por nominal	Phi	,211			,000
		V de Cramer	,149			,000
		Coeficiente de contingencia	,207			,000
	Ordinal por ordinal	Tau-b de Kendall	,176	,027	6,395	,000
		Tau-c de Kendall	,164	,026	6,395	,000
		Gamma Orden cero	,282	,043	6,395	,000
	N.º de casos válidos		1064			
Mujer	Nominal por nominal	Phi	,227			,000
		V de Cramer	,161			,000
		Coeficiente de contingencia	,221			,000
	Ordinal por ordinal	Tau-b de Kendall	,176	,028	6,304	,000
		Tau-c de Kendall	,163	,026	6,304	,000
		Gamma Orden cero	,283	,043	6,304	,000
	N.º de casos válidos		1028			
Total	Nominal por nominal	Phi	,214			,000
		V de Cramer	,151			,000
		Coeficiente de contingencia	,209			,000
	Ordinal por ordinal	Tau-b de Kendall	,177	,019	9,050	,000
		Tau-c de Kendall	,165	,018	9,050	,000
		Gamma Orden cero	,284	,030	9,050	,000
	N.º de casos válidos		2092			

a. Asumiendo la hipótesis alternativa.

b. Empleando el error típico asintótico basado en la hipótesis nula.

- Localice dos investigaciones (en una revista científica y una monografía). Compare la presentación que en ambas se hace de la investigación: estructura y redacción.

BIBLIOGRAFÍA CITADA

- Abad, F. J.; Olea, J.; Ponsoda, V. y García, C. (2011): *Medición en ciencias sociales y de la salud*. Síntesis. Madrid.
- Adorno, T.W.; Popper, K. R.; Dahrendorf, R.; Habermas, J.; Albert, H. y Pilot, H. (1973): *La disputa del positivismo en la sociología alemana*. Grijalbo. Barcelona.
- Alexander, C. (1986): *Writing for social scientists: how to start and finish your thesis, books or article*. Chicago University Press. Chicago.
- Allardt, E. (1990): "Challenges for comparative social research". *Acta Sociología*, vol. 33 C3): 183-193.
- Alós, J. S. (2001): "Industria de los estudios de mercado en España 2000". *Investigación y Marketing*, 72: 72-74.
- Alvira Martín, F. (1977): "La relación entre actitudes y conducta". *Revista Española de la Opinión Pública*, 49: 33-52.
- Alvira Martín, F. (1983): "Perspectiva cualitativa / perspectiva cuantitativa en la metodología sociológica". *Revista Internacional de Investigaciones Sociológicas*, 22: 53-75.
- Alvira Martín, F. (1989/2000): "Diseños de investigación social: criterios operativos", en García Ferrando, M.; Ibáñez, J. y Alvira Martín, F., *El análisis de la realidad social*. Alianza Universidad. Madrid: 85-111.
- Alvira Martín, F (1991a): *Metodología de la evaluación de programas*. CIS. Madrid. *Cuaderno Metodológico* nº 2.
- Alvira Martín, F (1991b): "Diseños de investigación", en Latiesa, M., *El pluralismo metodológico en la investigación social*. Universidad de Granada. Granada: 17-40.
- Alvira Martín, F (1992): "Reflexiones sobre la investigación sociológica", en Moya, C. y otros, *Escritos de teoría sociológica en homenaje a Luis Rodríguez Zúñiga*. CIS. Madrid.
- Alvira Martín, F. (1995): "Formas de presentación y representación gráfica de los datos" y "Organización del trabajo de investigación social". *Sociología: temas de materias sociales*. MAP. Consejo Superior de Administradores del Estado. Madrid: 244-275, 32-52.
- Alvira Martín, F. (2004/2011): *La encuesta: una perspectiva general metodológica*. CIS. Madrid. *Cuaderno Metodológico* nº 35.

- Andrews, F. (1984): "Construct validity and error components of survey measures". *Public Opinion Quarterly*, 48: 409-42.
- Aquilino, W. S. (1994): "Interview mode effects in surveys of drug and alcohol use. A field experiment". *Public Opinion Quarterly*, 58: 210-240.
- Aronson, E.; Brewer, M. y Carlsmith, J. M. (1985): "Experimentation in social psychology", en Lindzey, G. Aronson, E., *The handbook of social psychology*. Random House. Nueva York: 441-486.
- Azorín, F. y Sánchez Crespo, J. L. (1986): *Métodos y aplicaciones del muestreo*. Alianza. Madrid.
- Babbie, E. R. (1992): *The practice of social research*. Wadsworth Publishing Company. Belmont.
- Babbie, E. R. (2000): *Fundamentos de la investigación social*. Thomson editores. México.
- Bardin, L. (1986): *El análisis de contenido*. Akal. Madrid.
- Bassili, J. N. y Scott, B. S. (1996): "Response latency as a signal to question problems in survey research". *Public Opinion Quarterly*, 60: 390-399.
- Battaglia, M. P.; Link, M. W.; Frankel, M. R.; Osborn, L. y Mokdad, A. H. (2008): "An evaluation of respondent selection methods for household mail surveys". *Public Opinion Quarterly*, 72 (3): 459-469.
- Battaglia, M. P.; Krare, M.; Frankel, M. R.; Murray, M. C.; Bruckley, P. y Peritz, S. (2008): "Response rates: how have they changed and where are they headed?", en Lepkowski, J. M.; Tucker, C.; Brick, J. M.; De Leeuw, E. D.; Japac, L.; Lavrakas, P. J.; Link, M. W. y Sangster, R. L., *Advances in telephone survey methodology*. Wiley. Nueva York: 529-560.
- Bazeley, P. (2003): "Computerized data analysis for mixed methods research", en Tashakkori, A. y Teddlie, C., *Handbook of mixed methods in social and behavioral research*. Sage. Thousand Oaks, CA: 385-422.
- Bazeley, P. (2006): "The contribution of computer software to integrating qualitative and quantitative data analyses". *Research in the Schools*, 13 (1): 64-74.
- Becker, H. S. (1963): *Outsider: studies in the sociology of deviance*. The Free Press of Glencoe Inc. Nueva York.
- Becker, H. S. (1998): *Tricks of the trade. How to think about your research while you re doing it*. Chicago University Press. Chicago.
- Becker, H. S. y Geer, B. (1958): "Participant observation and interviewing. A rejoinder". *Human Organization*, 17: 36-40.
- Belson, W. (1981): *The design and understanding of survey questions*. Gower. Londres.
- Beltrán, M. (1985): "Cinco vías de acceso a la realidad social". *Revista Española de Investigaciones Sociológicas*, 29: 7-41.
- Beltrán, M. (1986/1989): "Cinco vías de acceso a la realidad social", en García Ferrando, M.; Ibáñez, J. y Alvira Martín, F., *El análisis de la realidad social*. Alianza Universidad. Madrid: 17-48.
- Berdie, D. R. y Anderson, J. F. (1974): *Questionnaires: design and use*. The Scarecrow Press. Metchuchen, N. J.
- Berelson, B. (1952): *Content analysis in communication research*. Free Press. Nueva York.
- Berger, P. y Luckmann, T. (1968): *La construcción social de la realidad*. Amorrortu. Buenos Aires.
- Berk, M. L.; Mathiowetz, N. A.; Ward, E. P. y White, A. A. (1987): "The effect of prepaid and promised incentives: results of a controlled experiment". *Journal of Official Statistics*, 3 (4): 449-457.

- Biemer, P. (2004): "Modeling measurement error to identify flawed questions", en Presser, S.; Rothgeb, J. M.; Couper, M. P.; Lessler, J. L.; Martin, E.; Martin, J. y Singer, E., *Methods for testing and evaluating survey questionnaires*. Wiley. Nueva York.
- Biemer, P. P. y Christ, S. L. (2008): "Weighting survey data", en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 317-341.
- Biemer, P. P. y Lyberg, L. E. (2003): *Introduction to survey quality*. John Wiley & Sons. New Jersey.
- Bishop, G. y Smith, A. (2001): "Response-order effects and the early Gallup split-ballots". *Public Opinion Quarterly*, 65: 479-505.
- Blaikie, N. (1991): "A critique of the use of triangulation in social research". *Quality and Quantity*, 25: 115-136.
- Blaikie, N. (2000): *Designing social research*. Blackwell. Oxford.
- Blaikie, N. (2010): *Designing social research*. Polity Press. Cambridge.
- Blalock, H. M. Jr. (1990): "Auxiliary measurement theories revisited", en Hox, J. J., y De Jong-Gierveld, J., *Operationalization and research strategy*. Swets and Zeitlinger. Amsterdam: 33-49.
- Blalock, H. M. y Blalock, A. B. (1968): *Methodology in social research*. McGraw Hill. Nueva York.
- Blumer, H. (1966): "Sociological implications of the work of G.H. Mead". *American Journal of Sociology*, 78 (1): 535-544.
- Blumer, H. (1969): *Symbolic interactionism: perspective and method*. Prentice-Hall, Inc. New Jersey.
- Bohm, R. M. (1990): "Death penalty opinions: a classroom experience and public commitment". *Sociological Inquiry*, 60: 285-297.
- Bohrnstedt, G. W. (1983): "Measurement", en Rossi, P. H.; Wrigth, J. D. y Anderson, A. B., *Handbook of survey research*. Academic Press Inc. Orlando: 70-122.
- Botella, J. y Gámbara, H. (2002): *Qué es el meta-análisis*. Biblioteca Nueva. Madrid.
- Bourque, L. B. y Clark, V. A. (1994): "Processing data: the survey example", en Lewis-Beck, M. S., *Research practice*. Sage. California: 1-88.
- Bradburn, N. M. (1983): "Response effects", en Rossi, P. H.; Wrigth, J. D. y Anderson, A. B., *Handbook of survey research*. Academic Press Inc. Orlando: 289-328.
- Bradburn, N. M.; Sudman, S. y colaboradores (1979): *Improving interview method and questionnaire design: response effects to threatening questions in survey research*. Jossey-Bass. San Francisco.
- Brannen, J. (1994): "Combining qualitative and quantitative approaches: an overview", en Brannen, J., *Mixing methods: qualitative and quantitative research*. Avebury. Aldershot: 3-38.
- Brewer, J. y Hunter, A. (1989): *Multimethod research: a synthesis of styles*. Sage. Londres.
- Brewer, J. y Hunter, A. (2006): *Foundations of multimethod research: synthesizing styles*. Sage. Thousand Oaks, CA.
- Bruyn, S.T. (1972): *La perspectiva humana en sociología*. Amorrortu. Buenos Aires.
- Bryant, C. G. (1975): "Kuhn, paradigms and sociology", *The British Journal of Sociology*. 26; 354-359.
- Bryman, A. (1988): *Quality and quantity in social research*. Unwin Hyman. Londres.
- Bryman, A. (1995): *Research methods and organization studies*. Routledge. Londres.
- Bryman, A. (2004): *Social research methods*. Oxford University Press. Oxford.
- Bryman, A. (2007): "The research question in social research: what is its role?" *International Journal of Social Research Methodology*, 10 (1): 5-20.

- Bryman, A. y Cramer, D. (1995): *Quantitative data analysis for social scientist*. Routledge. Londres.
- Bryman, A.; Becker, S. y Sempik, J. (2008): "Quality criteria for quantitative, qualitative and mixed methods research: a view from social policy". *International Journal of Social Research Methodology*, 11 (4): 261-276.
- Bulmer, M. (ed.) (1992): *Sociological research methods. An introduction*. Macmillan. Londres.
- Burgess, R. G. (1984): *In the field*. Allen & Unwin. Londres.
- Calvo Buezas, T. (2000): *Inmigración y racismo. Así sientes los jóvenes del siglo XXI*. Cauce Editorial. Madrid.
- Campanelli, P. (2008): "Testing survey questions", en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 176-200.
- Campbell, D. T. y Fiske, D. W. (1959): "Convergent and discriminant validity by the multitrait, multi-method matrix". *Psychological Bulletin*, 56: 81-105.
- Campbell, D. T. y Russo, M. J. (1999): *Social experimentation*. Sage. Thousand Oak, California.
- Campbell, D. T. y Stanley, J. (1963): *Experimental and quasi-experimental designs for research*. Rand McNally. Chicago.
- Campbell, D. T. y Stanley, J. (1970): *Diseños experimentales y cuasiexperimentales en la investigación social*. Amorrortu. Buenos Aires.
- Cannell, C. F.; Miller, P. V. y Oksenberg, L. (1981): "Research on interviewing techniques", en Leinhardt, S., *Sociological methodology*. Jossey-Bass. San Francisco: 389-437.
- Cantril, H. (1944): *Gauging public opinion*. Princeton University Press. Princeton.
- Caplow, T. (1977): *La investigación sociológica*. Laia. Barcelona.
- Carmines, E. G. y Zeller, R. A. (1979): *Reliability and validity assessment*. Sage. Beverly Hills.
- Catania, J.; Dinson, D.; Canahola, J.; Pollack, L.; Hauck, W. y Coates, T. (1996): "Effects of interviewer gender, interviewer choice and item wording on responses to questions concerning sexual behavior". *Public Opinion Quarterly*, 60: 345-375.
- Cea D'Ancona, M^a A. (1992): *La justicia de menores en España*. CIS/Siglo XXI. Madrid. Monografía n^o 127.
- Cea D'Ancona, M^a A. (1996/2009): *Metodología Cuantitativa. Estrategias y técnicas de investigación social*. Síntesis. Madrid.
- Cea D'Ancona, M^a A. (2002a/2008): *Análisis multivariable. Teoría y práctica en la investigación social*. Síntesis. Madrid.
- Cea D'Ancona, M^a A. (2002b): "La medición de las actitudes ante la inmigración: evaluación de los indicadores tradicionales de "racismo"". *Revista Española de Investigaciones Sociológicas*, 99: 87-111.
- Cea D'Ancona, M^a A. (2004): *La activación de la xenofobia en España. ¿Qué miden las encuestas?* CIS. Madrid. Monografía n^o 210.
- Cea D'Ancona, M^a A. (2005/2010): *Métodos de Encuesta. Teoría y práctica, errores y mejora*. Síntesis. Madrid.
- Cea D'Ancona, M^a A. (2009): "La compleja detección del racismo y la xenofobia a través de encuesta. Un paso adelante en su medición". *Revista Española de Investigaciones Sociológicas*, 125: 13-45.
- Cea D'Ancona, M^a Á. (2012): "Measuring xenophobia: social desirability and survey mode effects". *Journal of Ethnic and Migration Studies* (en prensa).
- Cea D'Ancona, M^a Á. y Valles Martínez, M. S. (2009): *Evolución del racismo y la xenofobia en España [Informe 2009]* OBERAXE. Ministerio de Trabajo e Inmigración. Madrid.

- Cea D'Ancona, M^a Á. y Valles Martínez, M. S. (2010a): *Xenofobias y xenofilias en clave biográfica: relatos entrelazados de autóctonos y foráneos*. Siglo XXI. Madrid.
- Cea D'Ancona, M^a Á. y Valles Martínez, M. S. (2010b): *Evolución del racismo y la xenofobia en España* [Informe 2010]. OBERAXE. Ministerio de Trabajo e Inmigración. Madrid.
- Cea D'Ancona, M^a Á. y Valles Martínez, M. S. (2011): *Evolución del racismo y la xenofobia en España* [Informe 2011]. OBERAXE. Ministerio de Trabajo e Inmigración. Madrid.
- Clairin, R. y Brion, P. (2001): *Manual de muestreo*. Hespérides. Madrid.
- Converse, J. M. y Presser, S. (1994): "Survey questions: handcrafting the standardized questionnaire", en Lewis-Beck, M. S., *Research Practice*. Sage. California: 89-162.
- Cook, T. D. (1983): "Quasi-experimentation: its ontology, epistemology, and methodology", en Morgan, G., *Beyond methods: strategies and social research*. Sage. Beverly Hills: 74-94.
- Cook, T. D. (1985): "Postpositivist critical multiplism", en Shortland, R. L. y Mark, M. M., *Social Science and social policy*. Sage. Thousand Oaks, California: 21-62.
- Cook, T. D. y Campbell, D.T. (1977): "The design and conduct of quasi-experiments and true experiments in field settings", en Dunnette, H. M., *Handbook of industrial psychology*. Rand McNally. Illinois.
- Cook, T. D. y Campbell, D. T. (1979): *Quasi-experimentation: design and analysis issues for field setting*. Houghton-Mifflin. Boston.
- Cook, T. D. y Reichardt, C. S. (eds.) (1979): *Qualitative and quantitative methods in evaluating research*. Sage. Beverly Hills, California.
- Cooper, H. M. (1984): *Integrating Research: A systematic approach*. Sage. Beverly Hills.
- Corbetta, P. (2003/2010): *Metodología y técnicas de investigación social*. McGraw Hill. Madrid.
- Couper, M. P. (2000): "Web surveys. A review of issues and approaches". *Public Opinion Quarterly*, 64: 464-94.
- Couper, M. P. y Hansen, S. E. (2002): "Computer-assisted interviewing", en Gubrium, J. F. y Holstein, J. A., *Handbook of interview research. Context & Methods*. Sage. Thousand Oaks, California: 557-575.
- Couper, M. P.; Traugott, M. W. y Lamias, M. J. (2001): "Web survey design and administration". *Public Opinion Quarterly*, 65: 230-53.
- Courgeau, D. (ed.) (2003): *Methodology and epistemology of multilevel analysis. Approaches from different social sciences*. Kluwer Academic Publishers. Dordrecht, Netherlands.
- Cresswell, J. W. (1994): *Research design: qualitative and quantitative approaches*. Sage. Thousand Oaks, California.
- Cresswell, J. W. (2007): *Qualitative inquiry and research design: choosing among five traditions*. Sage. Thousand Oaks, California.
- Cresswell, J. W. (2009): *Research design: qualitative, quantitative, and mixed methods*. Thousand Oaks, California.
- Cresswell, J. W. y Plano Clark, V. L. (2007): *Designing and conducting mixed methods research*. Thousand Oaks, California.
- Dale, A. (2006): "Quality issues with survey research". *International Journal of Social Research Methodology*, 9: 143-158.
- Davis, J. A. y Smith, T. W. (1992): *The NORC General Social Survey. A user's guide*. Sage. Newbury Park.
- De Leeuw, E. D. (2008): "Choosing the method of data collection", en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 113-135.

- De Leeuw, E. D. y Collins, M. (1997): "Data collection methods and survey quality: an overview", en Lyberg, L.; Biemer, P.; Collins, M.; De Leeuw, E.; Dippo, C.; Schwarz, N. y Trewin, D., *Survey measurement and process quality*. John Wiley & Sons Inc. Nueva York: 199-220.
- De Leeuw, E. D.; Hox, J. J. y Snijders, G. (2002): "The effect of computer-assisted interviewing on data quality: a review", en De Vaus, D., *Social surveys*. Sage. Londres: 170-191. (e. o. 1995: *Journal of the Market Research Society*, 37: 325-344).
- De Leeuw, E. D. y Hox, J. J. (2008): "Self-administered questionnaires: mail surveys and other applications", en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 239-263.
- De Leeuw, E. D.; Hox, J. J. y Dillman, D. A. (2008): "Mixed-mode surveys: when and why", en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 299-316.
- Della Porta, D. y Keating, M. (eds.) (2008): *Approaches and methodologies in the social sciences*. Cambridge University Press. Cambridge.
- De Miguel, A. (2000): *Dos generaciones de jóvenes (1960-1998)*. Instituto de la Juventud. Madrid.
- De Miguel, A. (2003): *El final del franquismo. Testimonio personal*. Marcial Pons. Madrid.
- De Miguel, J. (1999): "Cien años de investigación sociológica sobre España". *Revista Española de Investigaciones Sociológicas*, 87: 179-219.
- De Vaus, D. A. (1990): *Surveys in social research*. Unwin Hyman. Londres.
- De Vaus, D. A. (2001): *Research design in social research*. Sage. Londres.
- DeVellis, R. F. (1991): *Scale development. Theory and applications*. Sage. Newbury Park.
- Denzin, N. K. (1970): *The research act in sociology*. Butterworth. Londres.
- Denzin, N. K. (1975): *The research act*. Aldine. Chicago.
- Denzin, N. K. (1983): "Interpretative interactionism", en Morgan, G., *Beyond methods: strategies for social research*. Sage. Beverly Hills, CA: 129-46.
- Denzin, N. K. (1989): *The research act*. Prentice-Hall. Englewood Cliffs.
- Dex, S. (2002): "The reliability of recall data: a literature review", en De Vaus, D., *Social surveys*. Sage. Londres: 257-282 (e. o. 1995: *Bulletin de Methodologie Sociologique*, 49: 58-89).
- Díaz de Rada Igúzquiza, V. (2002): *Tipos de encuestas y diseños de investigación*. Universidad Pública de Navarra. Navarra.
- Díaz de Rada Igúzquiza, V. (2005): *Manual de trabajo de campo en la encuesta*. CIS. Madrid. Cuaderno Metodológico nº 36.
- Díaz de Rada Igúzquiza, V. (2009): *Análisis de datos de encuestas*. Editorial UOC. Barcelona.
- Dillman, D. A. (1978): *Mail and telephone surveys: the total design method*. John Wiley & Sons. Nueva York.
- Dillman, D. A. (1983): "Mail and other self-administered questionnaires", en Rossi, P. H.; Wrigth, J. D. y Anderson, A. B., *Handbook of survey research*. Academic Press Inc. Orlando: 359-378.
- Dillman, D. A. (1991): "The design and administration of mail surveys". *Annual Review of Sociology* 17: 225-49.
- Dillman, D. A. (2000): *Mail and internet surveys: the tailored design method*. Wiley. Nueva York.
- Dillman, D. A. (2008): "The logic and psychology of constructing questionnaires", en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 161-175.

- Dillman, D. A.; Gallegos, J. G. y Frey, J. H. (1976): "Reducing refusal rates for telephone interviews". *Public Opinion Quarterly*, 40: 66-78.
- Dillman, D. A.; Smyth, J. D. y Christian, L. M. (2009): *Internet, mail and mixed-mode surveys. The Tailored Design Method*. John Wiley & Sons, Inc. Hoboken. Nueva Jersey.
- Dillon, J. T. (1990): *The practice of questioning*. Routledge. Londres.
- Douglas, J. D. (1976): *Investigate social research*. Sage. Beverly Hills, CA.
- Duncan, O. D. (1984): *Notes on social measurement: historical and critical*. Russell Sage. Nueva York.
- Durkheim, E. (1974): *Las reglas del método sociológico*. Morata. Madrid.
- Easthope, G. (1974): *History of social research methods*. Longman. London.
- Eco, H. (1982): *Como se hace una tesis*. Gedisa. Barcelona.
- Ericsson, K. A. y Simon, H. (1980): "Verbal reports as data." *Psychological Review*, 87: 215-51.
- Erikson, K.T. (1962): "Notes on the sociology of deviance". *Social Problems*, 9: 307-314.
- Esposito, J. L. y Rothgeb, J. M. (1997): "Evaluating survey data: making the transition from pretesting to quality assessment", en Lyberg, L.; Biemer, P.; Collins, M.; De Leeuw, E.; Dippo, C.; Schwarz, N. y Trewin, D., *Survey measurement and process quality*. John Wiley & Sons Inc. Nueva York: 541-572.
- Fernández Prados, J. S. (1999): "Sistemas informatizados para las encuestas telefónicas". *Metodologías de Encuestas*, 1 (1): 29-46.
- Feyerabend, P.K. (1974): *Contra el método*. Ariel. Barcelona.
- Fielding, N. (2010): "Mixed methods in the real world". *International Journal of Social Research Methodology*, 13 (2): 127-138.
- Fielding, N. G. y Fielding, J. L. (1986): *Linking data*. Sage. Londres.
- Filstead, W.J. (1986): "Una experiencia necesaria en la investigación evaluativa", en Cook, T. D. y Reichart, Ch. S., *Métodos cualitativos y cuantitativos en investigación evaluativa*. Ediciones Morata. Madrid: 59-79.
- Fink, A. (1995): *How to design surveys*. Sage. Thousand Oaks, California.
- Fisher, R. A. (1925): *Statistical methods for research workers*. Oliver & Boyd. Edimburgo.
- Fisher, R. A. (1935): *The design of experiments*. Oliver & Boyd. Oxford.
- Fiske, D. W. (1971): *Measuring the concepts of personality*. Aldine. Chicago.
- Foster, G. G.; Hummel, R. L. y Adamchak, D. J. (1998): "Patterns of conception, natality, and mortality from midwestern cemeteries: a sociological analysis of historical data". *Sociological Quarterly*, 39: 473-490.
- Fowler, F. J. Jr. (1998): *Survey research methods*. Sage. Beverly Hills.
- Fowler, F. J. Jr. (2004): "The case for more split-sample experiments in developing survey instruments", en Presser, S.; Rothgeb, J. M.; Couper, M. P.; Lessler, J. L.; Martin, E.; Martin, J. y Singer, E., *Methods for testing and evaluating survey questionnaires*. Wiley. Nueva York.
- Fowler, F. J. Jr. y Cosenza, C. (2008): "Writing effective questions", en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 136-160.
- Frey, J. H. (1980/1989): *Survey research by telephone*. Sage. Newbury Park, California.
- Fricker, S.; Galesic, M.; Tourangeau, R. y Yan, T. (2005): "An experimental comparison of web and telephone surveys". *Public Opinion Quarterly*, 69 (3): 370-392.
- Fromm, E. (1973): *The anatomy of human destructiveness*. Penguin. Harmondsworth.
- Fuller, T. D. et al., (1993): "Using focus groups to adapt survey instrument to new populations", en Morgan, D. L., *Successful focus groups: advancing the state of the art*. Sage. Newbury Park: 89-104.

- Fundación FOESSA (1970): *Informe sociológico sobre la situación social de España*. Euroamérica. Madrid.
- García Ferrando, M. (1979): *Sobre el método. Filosofía analítica e investigación empírica*. CIS. Madrid.
- Garfinkel, H. (1967): *Studies in ethnomethodology*. Prentice-Hall. Englewood Cliffs, NJ.
- Garrigós Monerris, J. I. (2003): *Frédéric Le Play. Biografía intelectual, metodología e investigaciones sociológicas*. CIS. Madrid. Colección Monografías nº 203.
- Gilbert, G. H. y Doran, J. (1993): *Simulating societies: the computer simulation of social processes*. UCL Press. Londres.
- Gilbert, N. (1993): "Writing about social research", en Gilbert, N., *Research social life*. Sage. Londres: 329-344.
- Gilbert, N. y Troitzsch, K. G. (2006): *Simulación para las ciencias sociales. Una guía práctica para explorar cuestiones sociales mediante el uso de simulaciones informáticas*. McGraw Hill. Madrid.
- Glaser, B. G. y Strauss, A. L. (1967): *The discovery of grounded theory: strategies for qualitative research*. Aldine. Chicago, IL.
- Glass, G. V.; McGraw, B. y Smith, M. (1987): *Meta-analysis in social research*. Sage. Beverly Hills.
- Goldthorpe, J. E. (1982): *Introducción a la sociología* Alianza. Madrid.
- Goldthorpe, J. H. (2010): *De la sociología. Números, narrativa e integración de la investigación y la teoría*. CIS / BOE. Madrid.
- Gómez Martín, V. (2006): "El experimento (*das experiment*). Algunas reflexiones sobre la película y el *Stanford Prison Experiment* que la inspiró". *Revista Electrónica de Ciencia Penal y Criminología*, 08-06 (<http://criminet.ugr.es/recpc>)
- González Blasco, P. (1989/2000): "Medir en las ciencias sociales", en García Ferrando, M.; Ibáñez, J. y Alvira Martín, F., *El análisis de la realidad social*. Alianza Universidad. Madrid: 343-407.
- Goode, W. J. y Hatt, P. K. (1952): *Methods in social research*. McGraw Hill. Nueva York.
- Greene, J. C. (2007): *Mixed methods in social inquiry*. Jossey-Bass. Wiley. San Francisco, California.
- Greene, J. C.; Caracelli, V. J., y Graham, W. F. (1989): "Toward a conceptual framework for mixed-method evaluation designs". *Educational Evaluation and Policy Analysis*, 11 (3): 255-274.
- Greene, J.C. y Caracelli, V. J. (eds.) (1997): *Advances in mixed method evaluation: the challenges and benefits of integrating diverse paradigms*. Jossey-Bass. San Francisco, California.
- Groves, R. M. (1989): *Survey errors and survey costs*. John Wiley & Sons. Nueva York.
- Groves, R. M.; Fowler, F. J. Jr.; Couper, M. P.; Lepkowski, J. M.; Singer, E. y Tourangeau, R. (2009): *Survey methodology*. Wiley & Sons. Hoboken, New Jersey.
- Guba, E. G. y Lincoln, Y. S. (1988): "Do inquiry paradigms imply inquiry methodologies?", en Fetterman, D. M., *Qualitative Approaches to Evaluation in Education*. Praeger. Nueva York: 89-115.
- Guba, E. G. y Lincoln, Y. S. (1989): *Fourth generation evaluation*. Sage. Newbury Park, CA.
- Gullarhorn, J. T. y Gullarhorn, J. E. (1963) "A computer model of elementary social behavior", *Behavioral Science*, 8 (4): 354-92.
- Gwartney, P. A. (2007): *The Telephone Interviewer's Handbook: How to conduct standardized conversations*. Jossey-Bass. San Francisco, CA.
- Habermas, J. (1974): *Theory and practice*. Heinemann. Londres.
- Hakim, C. (1994): *Research design: strategies and choices in the design of social research*. Routledge. Londres.

- Hammersley, M. (1992): *What's wrong with ethnography?* Routledge. Londres.
- Hammersley, M. (1994): "Deconstructing the qualitative-quantitative divide", en Brannen, J., *Mixing methods: qualitative and quantitative research*. Aldershot. Avebury: 39-56.
- Hammersley, M. y Atkinson, P. (1983): *Ethnography: principles in practice*. Tavistock. Londres.
- Hammersley, M. y Atkinson, P. (2007): *Ethnography: principles in practice*. Routledge. Londres.
- Heath, A. y Martin, J. (1997): "Why are there so few formal measuring instruments in social and political research?", en Lyberg, L.; Biemer, P.; Collins, M.; De Leeuw, E.; Dippo, C.; Schwarz, N. y Trewin, D., *Survey measurement and process quality*. John Wiley & Sons Inc. Nueva York: 71-86.
- Heerwegh, D. (2009): "Mode differences between face-to-face and web surveys: An experimental investigation of data quality and social desirability effects". *International Journal of Public Opinion Research*, 21: 111-120.
- Heerwegh, D. y Loosveldt, G. (2008): "Face-to-face versus web surveying in a high-internet-coverage population differences in response quality". *Public Opinion Quarterly*, 72 (5): 836-846.
- Henerson, M. E., Morris, L. L. y Fitz-Gibbon, C. T. (1987): *How to measure attitudes*. Sage. Newbury Park, California.
- Hesselbart, S. (1975): "Education, ethnic, stereotypes and question format". *Sociology and Social Research*, 50: 266-273.
- Hesse-Biber, S. N. (2010): *Mixed methods research: merging theory with practice*. The Guilford Press. Nueva York.
- Hirchi, T. (1969): *Causes of delinquency*. University of California Press. California.
- Holbrook, A. L.; Green, M. C. y Krosnick, J. A. (2003): "Telephone versus face-to-face interviewing of national probability samples with long questionnaires". *Public Opinion Quarterly*, 67: 79-125.
- Holbrook, A. L.; Krosnick, J. A.; Moore, D. y Tourangeau, R. (2007): "Response order effects in dichotomous categorical questions presented orally. The impact of question and respondent". *Public Opinion Quarterly*, 71 (3): 325-348.
- Holbrook, A. L. y Krosnick, J. A. (2010): "Social desirability bias in voter turnout reports". *Public Opinion Quarterly*, 74 (1): 37-67.
- Holland, J. y Campbell, J. (eds.) (2005): *Methods in development research. Combining qualitative and quantitative approaches*. ITDG publishing. Warwickshire.
- Hox, J. J. (1997): "From theoretical concept to survey questions", en Lyberg, L.; Biemer, P.; Collins, M.; De Leeuw, E.; Dippo, C.; Schwarz, N. y Trewin, D., *Survey measurement and process quality*. John Wiley & Sons Inc. Nueva York: 47-70.
- Hox, J. J. y De Leeuw, E. D. (2002): "A comparison of nonresponse in mail, telephone and face to face surveys", en De Vaus, D., *Social surveys*. Sage. Londres: 157-171.
- Ibañez, J. (1985): *Del algoritmo al sujeto: perspectivas de la investigación social*. Siglo XXI. Madrid.
- Ibáñez, J. (1989): "Perspectivas de la investigación social: el diseño en la perspectiva estructural", en García Ferrando, M.; Ibáñez, J. y Alvira Martín, F., *El análisis de la realidad social*. Alianza Universidad. Madrid: 49-84.
- Ispízuza Uribarri, M. (2002): "La técnica de la encuesta ante las nuevas tecnologías. Unas primeras reflexiones", en Iglesias de Ussel, J. et al., *La sociedad: teoría e investigación empírica. Estudios en homenaje a José Jiménez Blanco*. CIS. Madrid: 1123-1140.
- Jacobs, H. (1994): "Using published data: errors and remedies", en Lewis-Beck, M. S., *Research practice*. Sage. California: 339-390.
- Jenkins, J. E. (2001): "Rural adolescent perceptions of alcohol and other drug resistance". *Child Study Journal*, 31 (4): 211-224.

- Jick, T. D. (1979): "Mixing qualitative and quantitative methods: triangulation in action". *Administrative Science Quarterly*, 24: 602-611.
- Johnson, R. B. y Onwuegbuzie, A. J. (2004): "Mixed methods research: a research paradigm whose time has come". *Educational Researcher* 33 (7): 14-26.
- Johnson, B. y Turner, L. A. (2003): "Data collection strategies in mixed methods research", en Tashakkori, A. y Teddlie, Ch., *Handbook of mixed methods in social & behavioral research*. Sage. Thousand Oaks, CA: 297-319.
- Jones, R.A. (1985): *Research methods in the social and behavioral sciences*. Sinauer Associates, Inc. Massachusetts.
- Jordan, L. A.; Marcus, A. C. y Reeder, L. G. (1980): "Response styles in telephone and household interviewing: a field experiment". *Public Opinion Quarterly*, 44: 210-222.
- Kaldenberg, D. O.; Koenig, H. F. y Becker, B. W. (1994): "Mail survey response rate patterns in a population of the elderly. Does response deteriorate with age? *Public Opinion Quarterly*, 58: 68-76.
- Kalton, G.; Roberts, J. y Holt, D. (1980): "The effects of offering a middle response option with opinion questions". *Statiscian*, 29: 65-78.
- Kalton, G. y Schuman, H. (2002): "The effect of the question on survey responses: a review", en De Vaus, D., *Social surveys*. Sage. Londres: 209-230. (e. o. 1982: *Journal of the Royal Statistical Society*, 145: 42-73).
- Kaplan, A. (1964): *The conduct of inquiry*. Chandler. San Francisco.
- Karakayali, N. (2009): "Social distance and affective Orientations". *Sociological Forum*, 23 (3): 538-562.
- Keeter, S. (1995): "Estimating telephone noncoverage bias with a telephone survey". *Public Opinion Quarterly*, 59: 196-217.
- Kemper, E. A.; Stringfield, S. y Teddlie, Ch. (2003): "Mixed methods sampling strategies in social science research", en Tashakkori, A. y Teddlie, Ch., *Handbook of mixed methods in social & behavioral research*. Sage: Thousand Oaks, CA: 273-296.
- Kerlinger, F. N. (1986): *Foundations of behavioral research*. Holt, Rinehart & Winston. Nueva York.
- Kiecolt, K. J. y Nathan, L. A. (1987): *Secondary analysis of survey data*. Sage. Beverly Hills.
- Kinsey, A. C.; Pemery, W. B. y Martin, C. E. (1948): *Sexual behavior in the human male*. Saunders. Philadelphia.
- Kish, L. (1949): "A procedure for objective respondent selection within the household". *Journal of the American Statistical Association*, 44: 380-387.
- Kish, L. (1965/1995): *Survey sampling*. John Wiley & Sons. Nueva York.
- Kitsuse, J. (1962): "Societal reaction to deviance: problems if theory and method". *Social Problems*, 9-4: 247-256.
- Kon, I. (ed.) (1989): *Historia de la sociología del s. XIX - comienzos del s. XX*. Editorial Progreso. Moscú.
- König, R. (1973): *Tratado de sociología empírica*. Tecnos. Madrid.
- Kovaliova, M. (1989): "Investigaciones sociales empíricas en el s.XIX", en Kon, I., *Historia de la sociología del s. XIX - comienzos del s. XX*. Editorial Progreso. Moscú: 121- 150.
- Kornhauser, A. (1951) "Constructing questionnaires and interview schedules", en Jahoda, M.; Deutsch, M. y Cook, S., *Research methods in social relations: Part two*, Dryden. Nueva York: 423-462.
- Kreuter, F.; Presser, S. y Tourangeau, R. (2008): "Social desirability bias in Cati, IVR, and Web Surveys. The effects of mode and question sensitivity". *Public Opinion Quarterly*, 72 (5): 847-865.

- Krippendorff, K. (1990): *Metodología del análisis de contenido: teoría y práctica*. Paidós. Barcelona.
- Krosnick, J. A. (1991): "Response strategies for coping with the cognitive demands of attitude measures in surveys". *Applied Cognitive Psychology*, 5: 213-236.
- Krosnick, J. A. y Alwin, D. F. (1987): "An evaluation of a cognitive theory of response-order effects in survey measurements". *Public Opinion Quarterly*, 51: 201-219.
- Krosnick, J. A. y Fabrigar, L. R. (1997): "Designing rating scales for effective measurement in surveys", en Lyberg, L.; Biemer, P.; Collins, M.; De Leeuw, E.; Dippo, C.; Schwarz, N. y Trewin, D., *Survey measurement and process quality*. John Wiley & Sons Inc. Nueva York: 141-164.
- Krosnick, J. A., y Presser, S. (2010): "Question and questionnaire design", en Marsden, P. V. y Wright, J. D., *Handbook of survey research*. Emerald Group Publishing. Bingley, UK: 263-314.
- Krysan, M. (1998): "Privacy and the expression of white racial attitudes. A comparison across three contexts". *Public Opinion Quarterly*, 62: 506-544.
- Kuhn, T. S. (1962): *The structure of scientific revolutions*. The University of Chicago Press. Chicago.
- Kuhn, T. S. (1971/1978): *La estructura de las revoluciones científicas*. Fondo de Cultura Económica. Madrid.
- Kuhn, T.S. (1978): *Segundos pensamientos sobre paradigmas*. Tecnos. Madrid.
- Lakatos, I. (1974): "Metodologías rivales de la ciencia: las construcciones racionales como guía de la historia". *Teorema*, IV (2): 199-214.
- Lakatos, I. (1975): "La falsación y la metodología de los programas de investigación científica", en Lakatos, I. y Musgrave, A., *Crítica y conocimiento*. Grijalbo. Barcelona.
- Lambers, J. (1980): *Psicología social*. Ediciones Pirámide, SA. Madrid.
- Lamo de Espinosa, E. (1992): "La sociología española desde 1939", en Ibáñez, J., *Las ciencias sociales en España: historia inmediata, crítica y perspectivas*. Universidad Complutense. Madrid: 119-130.
- Lavrakas, P. J. (1993): *Telephone survey methods: sampling, selection and supervision*. Sage. Newbury Park, CA.
- Lavrakas, P. J. (1998): "Methods of sampling and interviewing in telephone surveys", en Bickman, L. y Rog, D. J., *Handbook of applied social research methods*. Sage. Thousand Oaks, CA: 429-472.
- Layder, D. (1998): *Sociological practice: linking theory and social research*. Sage. Londres.
- Lazarsfeld, P. F., Berelson, B. y Gaudet, H. (1944): *The people's choice: how the voter makes up his mind in a presidential campaign*. Columbia University Press. Nueva York.
- Lazarsfeld, P. F. (1958): "Evidence and inference in social research". *Daedalus*, 87: 99-130.
- Lazarsfeld, P. F. (1971): "Una introducción conceptual al análisis de la estructura latente", en Mora y Araujo, M., *Medición y construcción de índices*. Nueva Visión. Argentina: 29-82.
- Lazarsfeld, P. F. (1973a): "De los conceptos a los índices empíricos", en Boudin, R. y Lazarsfeld, P., *Metodología de las ciencias sociales. I Conceptos e índices*. Laia. Barcelona: 35-46.
- Lazarsfeld, P. F. (1973b): "La sociología", en Piaget, J. et alía, *Tendencias de la investigación en las ciencias sociales*. Alianza. Madrid.
- Lazarsfeld, P. F. y Oberschall, A. (1965): "Max Weber and empirical social research". *American Sociological Review*, 30 (2): 185-199.
- Lee, R. M. (2000): *Unobtrusive methods in social research*. Open University Press. Londres.
- Lemert, E. M. (1969): "Primary and secondary deviation", en Cressey, D.R. y Ward, D.W., *Delinquency, crime and social process*. Harper & Row. Nueva York: 603-607.

- Lensvelt-Mulders, G. (2008): "Surveying sensitive topics", en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 461-478.
- Lepkowski, J. M.; Tucker, C.; Brick, J. M.; De Leeuw, E. D.; Japac, L.; Lavrakas, P. J.; Link, M. W. y Sangster, R. L., (2008): *Advances in telephone survey methodology*. Wiley. Nueva York.
- Lessler, J. T.; Eyerman, J. y Wang, K. (2008): "Interviewer training", en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 442-460.
- Lévy Manguin, J. P. y Varela, M. (eds.) (2003): *Análisis multivariable para las ciencias sociales*. Pearson Education. Madrid.
- Lin, N. (1976): *Foundations of social research*. McGraw Hill. Nueva York.
- Lisón, C. (1968): "Una gran encuesta de 1901-1902". *Revista Española de la Opinión Pública*, 12: 83-152.
- Litwin, M. S. (1995): *How to measure survey reliability and validity*. Sage. Thousand Oaks, CA.
- Lockwood, D. (1956): "Some remarks on "the social system"". *British Journal of Sociology*, 7: 134-146.
- Loftus, E. (1984): "Protocol analysis of responses to survey recall questions", en Jabine, T.; Straf, M.; Tanur, J. y Tourangeau, R., *Cognitive aspects of survey methodology: building a bridge between disciplines*. National Academy Press. Washington, DC: 61-64.
- Lohr, S. L. (2000): *Muestreo: diseño y análisis*. International Thomson Editores. México.
- Lohr, S. L. (2008): "Coverage and sampling", en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 97-112.
- López-Aranguren, E. (1989/2000): "El análisis de contenido", en García Ferrando, M.; Ibáñez, J. y Alvira Martín, F., *El análisis de la realidad social*. Alianza Universidad. Madrid: 383-414.
- López Pintor, R. (1982): *La opinión pública española del franquismo a la democracia*. Centro de Investigaciones Sociológicas. Madrid.
- Lozar Manfreda, K. (2008): "Internet surveys", en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 264-284.
- Lyberg, L. E. y Biemer, P. P. (2008): "Quality assurance and quality control in surveys", en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 421-441.
- Lyberg, L. y Kasprzyk, D. (1997): "Some aspect of post-survey processing", en Lyberg, L.; Biemer, P.; Collins, M.; De Leeuw, E.; Dippo, C.; Schwarz, N. y Trewin, D., *Survey measurement and process quality*. John Wiley & Sons Inc. Nueva York: 353-370.
- Lynn, P. (2008): "The problem of nonresponse", en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 35-55.
- Lynn, P. y Lievesley, D. (1991): *Drawing general population samples in Great Britain*. Social & Company Planning Research. Londres.
- Lyons Morris, L.; Taylor Fitz-Gibbon, C.; Lindheim, E. (1987): *How to measure performance and use tests*. Sage. Newbury Park, CA.
- MacDonald, K. y Tipton, C. (1993): "Using documents", en Gilbert, N., *Researching social life*. Sage. Londres: 187-200.
- Madey, D. (1982): "Some benefits of integrating qualitative and quantitative evaluation". *Educational Evaluation and Policy Analysis*, 4: 223-236.

- Mangione, T. W. (1998): "Mail surveys", en Bickman, L. y Rog, D. J., *Handbook of applied social research methods*. Sage. Thousand Oaks, California: 399-428.
- Mann, Ch. y Stewart, F. (2002): "Internet interviewing", en Gubrium, J. F. y Holstein, J. A., *Handbook of interview research. Context & Methods*. Sage. Thousand Oaks, CA: 603-627.
- Markoff, J.; Gilbert, S. y Weitman, S. R. (1974): "Toward the integration of content analysis and general methodology", en Heise, D., *Sociological methodology*. Jossey-Bass. San Francisco: 1-58.
- Marshall, C. y Rossman, B. (1989): *Design qualitative research*. Sage. California.
- Martin, E. (2004): "Vignettes and respondent debriefing for questionnaire design and evaluation", en Presser, S.; Rothgeb, J. M.; Couper, M. P.; Lessler, J. L.; Martin, E.; Martin, J. y Singer, E., *Methods for testing and evaluating survey questionnaires*. Wiley. Nueva York.
- Mason, R.; Carlson, J. E. y Tourangeau, R. (1994): "Contrast effects and subtraction in part-whole questions". *Public Opinion Quarterly*, 58: 569-578.
- Massey, J. T. (1988): "An overview of telephone coverage", en Groves, R. M.; Biemer, P. P.; Lyberg, L. E.; Massey, J. T.; Nicholls II, W. L. y Waksberg, J., *Telephone survey methodology*. John Wiley & Sons. Nueva York: 3-8.
- Maxwell, J. A. (1992): "Understanding and validity in qualitative research". *Harvard Educational Review*, 62: 279-300.
- Maxwell, J. A. (1996): *Qualitative research design*. Sage. Thousand Oaks, CA.
- Maxwell, J. A. y Loomis, D. M. (2003): "Mixed methods design: an alternative approach", en Tashakkori, A. y Teddlie, Ch., *Handbook of mixed methods in social & behavioral research*. Sage: Thousand Oaks, CA (p. 241-271).
- May, T. (2001): *Social Research. Issues, methods and process*. Open University Press. Buckingham.
- McCall, G. J. y Simmons, J. L. (1969): *Issues in participant observation: a text and a reader*. Addison-Wesley. Reading.
- McQuay, H. C. (1979): "The three faces of evaluation. What can be expected to work?", en Sechrest, L. et al., *Evaluation studies. Review Annual*, 4: 96-109.
- McClendon, M. J. (1986): "Response-order effects for dichotomous questions". *Social Science Quarterly*, 67: 205-211.
- Merton, R. K. (1967): *On theoretical sociology*. Free Press. Nueva York.
- Merton, R. K. (1983): *Teoría y estructuras sociales*. FCE. México.
- Merton, R. K.; Coleman, J. S. y Rossi, P. H. (eds.) (1979): *Qualitative and quantitative social research. Papers in honor of Paul F. Lazarsfeld*. The Free Press. Nueva York.
- Merton, R. K. y Lazarsfeld, P. F. (eds.) (1950): *Continuities in social research: studies in the scope and method of "The American Soldier"*. Glencoe Il., Free Press. Nueva York.
- Miller, S. y Conroy, B. (2008): "IVR: Interactive Voice Response", en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 285-298.
- Mitchell, M. y Jolley, J. (2007): *Research design explained*. Harcourt Brace Jovanovich. Nueva York.
- Molenaar, N. J. (1982): "Response-effects of "format" characteristics of question", en Dijkstra, W. y van der Zouwen, J., *Response behaviour in the survey-interview*. Academic Press. Nueva York: 49-90.
- Morales, J. F. et al. (1981): *Metodología y teoría de la psicología*. UNED. Madrid.
- Morales, J. F.; Huici, C.; Gómez, Á. y Gaviria, E. (eds.) (2008): *Método, teoría e investigación en psicología social*. Pearson Prentice Hall. Madrid.
- Moser, C. y Kalton, G. (1971): *Survey methods in social investigation*. Heinemann. Londres.

- Neuman, W. L. (2000): *Social research methods: qualitative and quantitative approaches*. Allyn and Bacon. Boston, MA.
- Neuman, W. L. (2007): *Basics of social research: qualitative and quantitative approaches*. Allyn and Bacon. Boston, MA.
- Newman, I.; Ridenour, C. S.; Newman, C. y De Marco, G. M. P. Jr (2003): "A typology of research purposes and its relationship to mixed methods", en Tashakkori, A. y Teddlie, Ch., *Handbook of mixed methods in social & behavioral research*. Sage. Thousand Oaks, CA: 167-188.
- Nicholls II, W. L.; Baker, R. P. y Martin, J. (1997): "The effect of new data collection technologies on survey data quality", en Lyberg, L.; Biemer, P.; Collins, M.; De Leeuw, E.; Dippo, C.; Schwarz, N. y Trewin, D., *Survey measurement and process quality*. John Wiley & Sons Inc. Nueva York: 221-248.
- Nisbet, R. M. (1966): *The sociological tradition*. Basic Books. Nueva York.
- O'Connell, J. y Layder, D. (1994): *Methods, sex and madness*. Routledge. Londres.
- O'Muircheartaigh, C. (1997): "Measurement error in surveys: a historical perspective", en Lyberg, L.; Biemer, P.; Collins, M.; De Leeuw, E.; Dippo, C.; Schwarz, N. y Trewin, D., *Survey measurement and process quality*. John Wiley & Sons Inc. Nueva York: 1-28.
- Oppenheim, A. (1966): *Questionnaire design and attitude measurement*. Heinemann. Londres.
- Orenstein, A. y Phillips, W. R. (1978): *Understanding social research: an introduction*. Allyn and Bacon, Inc. Boston.
- Orne, M. T. (1962): "On the social psychology of the psychological experiment with particular reference to demand characteristics and the implications". *American Psychologist*, 17: 776-783.
- Orr, L. L. (1999): *Social experiments: evaluating public programs with experimental methods*. Sage. Thousand Oaks, CA.
- Ortí, A. (1989/2000): "La apertura y el enfoque cualitativo o estructural: la entrevista abierta y la discusión de grupo", en García Ferrando, M.; Ibáñez, J. y Alvira Martín, F., *El análisis de la realidad social*. Alianza Universidad. Madrid: 171-204.
- Oskamp, S. (1977): "Methods of studying social behaviour", en Wrightsman, L. S., *Social psychology*. Brooks/Cole. Monterey.
- Pardo Merino, A. y Ruiz Díaz, M. A. (2005): *Análisis de Datos con SPSS 13 Base*. McGraw Hill. Madrid.
- Parsons, T. (1951): *The social system*. The Free Press. Nueva York.
- Parsons, T. (1982): *El sistema social*. Alianza. Madrid.
- Patton, M. Q. (1980): *Program evaluation kit. Vol.4. How to use qualitative methods in evaluation*. Sage. Londres.
- Patton, M. Q. (2002): *Qualitative research and evaluation methods*. Sage. Thousand Oaks, CA.
- Payne, S. L. (1949): "Case study in question complexity". *Public Opinion Quarterly*, 13: 653-658.
- Payne, S. L. (1980): *The art of asking questions*. Princeton University Press. Princeton. Nueva Jersey (e.o. 1951).
- Perelló Oliver, S. (2009): *Metodología de la investigación social*. Dykinson, SL. Madrid.
- Pérez Agote, A. (1989): *La sociedad y lo social. Ensayos de sociología*. Editorial Ellacuria. Vizcaya.
- Pérez Yruela, M. y Desrues, T. (2007): *Opinión de los españoles en materia de racismo y xenofobia*. Ministerio de Trabajo y Asuntos Sociales. OBERAXE. Madrid.

- Peytchev, A.; Couper, M. P.; McCabe, S. E. y Crawford, S. D. (2006): "Web survey design. Paging versus scrolling". *Public Opinion Quarterly*, 70: 596-607.
- Phillips, B. (1985): *Sociological research methods: an introduction*. The Dorsey Press. Illinois.
- Phillips, D. L. (1973): *Abandoning method*. Jossey-Bass. San Francisco.
- Platt, J. (1996): *A history of sociological research methods in America 1920-1960*. Cambridge University Press. Cambridge.
- Popper, K. R. (1967): *El desarrollo del conocimiento científico. Conjeturas y refutaciones*. Paidós. Buenos Aires.
- Popper, K. R. (1977): *La lógica de la investigación científica*. Tecnos. Madrid.
- Presser, S.; Couper, M. P.; Lessler, J. T.; Martin, E.; Martin, J.; Rothgeb, J. M. y Singer, E. (2004): "Methods for testing and evaluating survey questions". *Public Opinion Quarterly*, 68 (1): 109-130.
- Presser, S.; Rothgeb, J. M.; Couper, M. P.; Lessler, J. T.; Martin, E.; Martin, J. y Singer, E. (2004): *Methods for testing and evaluating survey questionnaires*. John Wiley & Sons, Inc. Hoboken. Nueva Jersey.
- Presser, S. y Stinson, L. (1998): "Data collection and social desirability bias in self-reported religious attendance". *American Sociological Review*, 63: 137-145.
- Procter, M. (1993): "Measuring attitudes", "Analysis survey data" y "Analysing other researchers 'data'", en Gilbert, N., *Researching social life*. Sage. Londres: 116-134, 239-254, 255-269.
- Reeve, B. y Mâsse, L. (2004): "Item response theory (IRT) modeling for questionnaire evaluation", en Presser, S.; Rothgeb, J. M.; Couper, M. P.; Lessler, J. T.; Martin, E.; Martin, J. y Singer, E. (2004): *Methods for testing and evaluating survey questionnaires*. John Wiley & Sons, Inc. Hoboken. Nueva Jersey.
- Reher, D. y Valero A. (1995): *Fuentes de Información demográfica en España*. CIS. Madrid. *Cuaderno Metodológico* nº 13.
- Reichardt, C. S. y Cook, T. D. (1979): "Beyond qualitative versus quantitative methods", en Cook, T. D. y Reichardt, Ch., *Qualitative and quantitative methods in social research*. Sage. Beverly Hills, CA: 7-32.
- Reichardt, C. S. y Rallis, S. F. (eds.) (1994): *The qualitative – quantitative debate: new perspectives*. Jossey-Bass. San Francisco.
- Reichenbach, H. (1973): *La filosofía científica*. FCE. México.
- Ritchey, F. J. (2008): *Estadística para las ciencias sociales*. McGraw-Hill. México.
- Ritzer, G. (1993): *Teoría sociológica contemporánea*. McGraw-Hill. Madrid.
- Robert, R. (1966): *Experimenter effects in behavioral research*. Appleton Century-Crofts. Nueva York.
- Rodríguez Ibáñez, J. E. (1989): *La perspectiva sociológica: Historia, teoría y método*. Taurus. Madrid.
- Rodríguez Osuna, J. (1991): *Métodos de muestreo*. CIS. Madrid. *Cuaderno Metodológico* nº 1.
- Rodríguez Osuna, J. (1993): *Métodos de muestreo. Casos prácticos*. CIS. Madrid. *Cuaderno Metodológico* nº 6. Madrid.
- Roethlisberger, F. J. y Dickenson, W. J. (1939): *Management and the worker*. Harvard University Press. Cambridge, MA.
- Rose, A. M. (1962): *Human behavior and social process*. Houton & Mifflin. Boston.
- Rossi, P. H.; Wrigth, J. D. y Anderson, A. B. (1983): "Sample surveys: history, current practice and future prospects", en Rossi, P. H.; Wrigth, J. D. y Anderson, A. B., *Handbook of survey research*. Academic Press Inc. Orlando: 1-20.

- Rossmann, G. B. y Wilson, B. L. (1985): "Numbers and words: combining quantitative and qualitative methods in a single large-scale evaluation study". *Evaluation Research* 9 (5): 627-43.
- Rubio, M^a J. y Varas, J. (2004): *El análisis de la realidad, en la intervención social. Métodos y técnicas de investigación*. Editorial CCS. Madrid.
- Salmon, C. T. y Nicholls, J. S. (1983): "The next-birthday method for respondent selection". *Public Opinion Quarterly*, 47: 270-276.
- Sánchez Carrión, J. J. (1985): "Técnicas de análisis de textos mediante codificación manual". *Revista Internacional de Sociología*, 43 (1): 89-118.
- Sánchez Carrión, J. J. (1995/1999): *Manual de análisis de datos*. Alianza. Madrid.
- Sánchez Carrión, J. J. (2000): *La bondad de la encuesta: el caso de la no respuesta*. Alianza. Madrid.
- Sánchez Crespo, J. L. (1971): *Principios elementales de muestreo y estimación de proporciones*. INE. Madrid.
- Sánchez-Cuenca, I. (2004/2009) *Teoría de juegos*. Madrid. CIS. *Cuaderno Metodológico*, nº 34.
- Saris, W.; van der Veld, W. y Gallhofer, I. (2004): "Development and Improvement of Questionnaires Using Predictions of Reliability and Validity", en Presser, S.; Rothgeb, J. M.; Couper, M. P.; Lessler, J. L.; Martin, E.; Martin, J. y Singer, E., *Methods for testing and evaluating survey questionnaires*. Wiley. Nueva York.
- Saris, W. E. y Gallhofer, I. N. (2007): *Design, evaluation, and analysis of questionnaires for survey research*. John Wiley & Sons Hoboken. New Jersey.
- Sayer, A. (2000): *Realism and social science*. Sage. Thousand Oaks, CA.
- Saxe, L y Fine, M. (1981): *Social experiments. Methods for design and evaluation*. Sage. Beverly Hills, CA.
- Scheuch, E. K. (1990): "The development of comparative research: towards causal explanation", en Oyén, E., *Comparative methodology*. Sage. Newbury Park, CA: 19-37.
- Schofield, J. W. (1993): "Increasing the generalizability of qualitative research", en Hammersley, M., *Social Research: philosophy, politics and practice*. Sage. Londres: 200-25.
- Schuman, H. y Presser, S. (1979a): "The assessment of "no opinion" in attitude surveys", en Schuessler, K. (comp.): *Sociological methodology*, Jossey-Bass. San Francisco.
- Schuman, H. y Presser, S. (1979b): "The open and closed question". *American Sociological Review*, 44: 692-712.
- Schuman, H. y Presser, S. (1981): *Questions & answers in attitude surveys. Experiments on question form, wording and context*. Academic Press. Nueva York.
- Schuman, H. y Presser, S. (1996): *Questions & answers in attitude surveys. Experiments on question form, wording and context*. Sage. Thousand Oaks, CA.
- Schur, E. (1968): *Law and society in sociological view*. Random House. Nueva York.
- Schwartz, H. y Jacobs, J. (1979): *Qualitative sociology: a method to the madness*. Free Press. Nueva York.
- Schwarz, N.; Strack, F.; Hippler, H. J. y Bishop, G. (1991): "The impact of administration mode on response effects in survey measurement". *Applied Cognitive Psychology*, 5: 193-212.
- Schwarz, N.; Knäuper, B.; Oyserman, D. y Stich, Ch. (2008): "The psychology of asking questions", en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 18-34.
- Shaw, C. y McKay, H. (1969): *Juvenile delinquency and urban areas*. University of Chicago Press. Chicago.
- Sheatsley, P. B. (1983): "Questionnaire construction", en Rossi, P. H.; Wright, J. D. y Anderson, A. B., *Handbook of survey research*. Academic Press Inc., Orlando: 195-230.

- Sieber, S. D. (1973): "The integration of fieldwork and survey methods". *American Sociological Review* 78 (6): 1335-1359.
- Siegel, S. (1985): *Estadística no paramétrica*. Trillas. México.
- Sikkel, D. y Hoogendoorn, A. (2008): "Panel surveys", en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 479-99.
- Silverman, D. (1985): *Quality methodology and sociology: describing the social work*. Alderson Gower. Londres.
- Simon, J. (1978): *Basic research methods in social sciences*. Random House. Nueva York.
- Singer, E. y Presser, S. (2008): "Privacy, confidentiality, and respondent burden as factor in telephone survey nonresponse", en Lepkowski, J. M.; Tucker, C.; Brick, J. M.; de Leeuw, E. D.; Japac, L.; Lavrakas, P. J.; Link, M. W. y Sangster, R. L., *Advances in telephone survey methodology*. Willey. Nueva York: 449-470.
- Singer, E.; Van Hoewyk, J. y Maher, M. P. (2000): "Experiments with incentives in telephone surveys". *Public Opinion Quarterly*, 64: 171-188.
- Singer, E.; Von Thurn, D. R. y Miller, E. R. (1995): "Confidentiality assurances and response: a quantitative review of the experimental literature". *Public Opinion Quarterly*, 59: 66-77.
- Singleton, R. A. Jr. y Straits, B. C. (2002): "Survey interviewing", en Gubrium, J. F. y Holstein, J. A., *Handbook of interview research. Context & Methods*. Sage. Thousand Oaks, CA: 59-82.
- Singleton, R. Jr.; Straits, B.; Straits, M. y McAllister, R. (1988): *Approaches to social research*. Oxford University Press. Nueva York.
- Sletto, R. (1950): "Pretesting of questionnaires". *American Sociological Review*, 5:193-200.
- Smith, J. K. (1983): "Quantitative versus qualitative research: an attempt to clarify the issue". *Educational Researcher* 12 (3): 6-13.
- Smith, M. L. y Glass, G. V. (1987): *Research and evaluation in education and the social sciences*. Prentice-Hall. Englewood Cliffs, NJ.
- Smith, T. W. y Stephenson, C. B. (1979): "An analysis of test-retest experiments in the 1972, 1973, 1974 and 1978 General Social Surveys". *General Social Survey Technical Report*, n° 14. NORC. Chicago.
- Spector, P. E. (1992): *Summated rating scale construction*. Sage. Newbury Park, California.
- Stake, R. E. (2005): "Qualitative case studies", en Denzin, N. K. y Lincoln, Y. S., *The sage handbook of qualitative research*. Sage. Thousand Oaks, CA: 443-66.
- Stapleton, L. M. (2008): "Analysis of data from complex surveys", en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 342-369.
- Steeh, Ch. (2008): "Telephone surveys", en De Leeuw, E. D.; Hox, J. J. y Dillman, D. A., *International handbook of survey methodology*. Lawrence Erlbaum Associates. Taylor & Francis Group. Nueva York: 221-238.
- Stevens, S. S. (1951): "Mathematics, measurement and psychophysics", en Stevens, S. S., *Handbook of experimental psychology*. John Wiley. Nueva York: 1-49.
- Stewart, D. M. (1984): *Secondary research: information sources and methods*. Sage. Beverly Hills, CA.
- Stoop, I; Billiet, J.; Koch, A. y Fitzgerald, R. (2010): *Improving survey response. Lessons learned from the European Social Survey*. John Wiley & Sons, Ltd., Sussex, U.K.
- Strauss, A. y Corbin, J. M. (1998): *Basics of qualitative research: techniques and procedures for developing grounded theory*. Sage. Thousand Oaks, CA.

- Sudman, S. (1976): *Applied sampling*. Academic Press. Nueva York.
- Sudman, S. (1983): "Applied sampling", en Rossi, P. H.; Wrigth, J. D. y Anderson, A. B., *Handbook of survey research*. Academic Press Inc. Orlando: 145-194.
- Sudman S. y Bradburn, N. M. (1974): *Response effects on surveys: a review and synthesis*. Aldine. Chicago.
- Sudman, S. y Bradburn, N. M. (1982/1987): *Asking questions: a practical guide to questionnaire design*. Jossey-Bass. San Francisco.
- Sudman, S.; Bradburn, N. M. y Schwarz, N. (1996): *Thinking about answers: the applications of cognitive processes to survey methodology*. Jossey-Bass. San Francisco.
- Tashakkori, A. y Teddlie, Ch. (1998): *Mixed methodology: combining qualitative and quantitative approaches*. Sage. Thousand Oaks, CA.
- Tashakkori, A. y Teddlie, Ch. (eds.) (2003): *Handbook of mixed methods in social & behavioral research*. Sage: Thousand Oaks, CA.
- Teddlie, Ch. y Tashakkori, A. (2003): "Major issues and controversies in the use of mixed methods in the social and behavioral sciences", en Tashakkori, A. y Teddlie, Ch., *Handbook of mixed methods in social & behavioral research*. Sage. Thousand Oaks, CA: 3-50.
- Tesch, R. (1992): *Qualitative research: analysis types and software tools*. The Falmer Press. Nueva York.
- Toepoel, V.; Das, M. y Van Soest, A. (2008): "Effects of design in web surveys. Comparing trained and fresh respondents". *Public Opinion Quarterly*, 72 (5): 985-1007.
- Toepoel, V.; Das, M. y Van Soest, A. (2009): "Design of web questionnaires: the effects of the number of item per screen". Forthcoming in *Field Methods* 21 (2).
- Tourangeau, R. (2004): "Experimental design considerations for testing and evaluating questionnaires", en Presser, S.; Rothgeb, J. M.; Couper, M. P.; Lessler, J. L.; Martin, E.; Martin, J. y Singer, E., *Methods for testing and evaluating survey questionnaires*. Wiley. Nueva York.
- Tourangeau, R.; Couper, M. P. y Conrad, F. (2004): "Spacing, position, and order. interpretative heuristic for visual features of survey questions". *Public Opinion Quarterly*, 68: 368-93.
- Tourangeau, R. y Rasinski, K. A. (1988): "Cognitive processes underlying context effects in measurement". *Psychological Bulletin*, 103: 299-314.
- Tourangeau, R.; Rips, L. J. y Rasinski, K. (2000): *The psychology of survey response*. Cambridge University Press. Cambridge.
- Tourangeau, R. y Smith, T. W. (1996): "Asking sensitive questions: impact of data collection mode, question format and question context". *Public Opinion Quarterly*, 60: 275-304.
- Tourangeau, R. y Smith, T. W. (1998): "Collecting sensitive information with different modes of data collection", en Couper, M. P.; Baker, R. P.; Bethlehem, J.; Clark, C. Z. F.; Martín, J. y Nicholls II, W. L., *Computer assisted survey information collection*. Wiley. New York: 431-453.
- Tourangeau, R. y Yan, T. (2007): "Sensitive questions in surveys". *Psychological Bulletin*, 133: 859-83.
- Troldahl, V. C. y Carter, R. E. Jr. (1964): "Random selection of respondents within households in phone surveys". *Journal of Marketing Research*, 1: 71-76.
- Trow, M. (1957): "Comment on participant observation and interviewing. A comparison". *Human Organization*, 16: 33-35.
- Tucker, C. y Lepkowski, J. M. (2008): "Telephone survey methods: adapting to change", en Lepkowski, J. M.; Tucker, C.; Brick, J. M.; de Leeuw, E. D.; Japec, L.; Lavrakas, P. J.; Link, M. W. y Sangster, R. L., *Advances in telephone survey methodology*. Wiley. Nueva York: 3-26.
- Turner, J. H. (1985): "In defense of positivism". *Sociological Theory*, 3: 24-30.

- Turner, R. H. (1962): "Role-making process versus conforming", en Rose, A.M., *Human behavior and social process*. Houton & Mifflin. Boston.
- Tyson, K. (1995): *New foundations for scientific social and behavioral research. The heuristic paradigm*. Allyn and Bacon. Boston.
- Valles, M. S. (1997/2009): *Técnicas cualitativas de investigación social. Reflexión metodológica y práctica profesional*. Síntesis. Madrid.
- Vidich, A. J. y Shapiro, G. (1955): "A comparison of participant observation and survey data". *American Sociological Review*, 20: 28-33.
- Wahlke, J. C. (1979): "Pre-behavioralism in political science". *The American Political Science Review*, 73: 9-31.
- Walker, R. (1988): "An introduction to applied qualitative research", en Walker, R., *Applied qualitative research*. Gower Hants. Londres.
- Wallace, W. L. (1971): *The logic of science in sociology*. Aldine-Atherton. Chicago.
- Wallace, W. L. (1976): *La lógica de la ciencia en la sociología*. Alianza. Madrid.
- Warner, S. L. (1965): "Randomized response: a technique for eliminating evasive response bias". *Journal of the American Statistical Association*, 69: 886-893.
- Webb, E. J.; Campbell, D. T.; Schwartz, R. D. y Sechrest, L. (1966): *Unobtrusive measures: non-reactive research in the social sciences*. Rand McNally. Chicago, IL.
- Webb, E. J.; Campbell, D. T.; Schwartz, R. D.; Sechrest, L. y Grove, J. B. (1981): *Nonreactive measures in the social sciences*. Houghton Mifflin. Dallas, TX.
- Weber, M. (1973): *Ensayos sobre metodología sociológica*. Amorrortu. Buenos Aires.
- Weber, M. (1977): *Sobre la teoría de las relaciones sociales*. Península. Barcelona.
- Weber, M. (1998): *El político y el científico*. Alianza. Madrid.
- Weber, R. P. (1994): "Basic content analysis", en Lewis-Beck, M. S., *Research practice*. Sage. California: 251-338.
- Weiss, S. et al. (2005): *Text mining: predictive methods for analyzing unstructured information*. Springer. Berlín.
- Wells, R. H. y Picou, J. S. (1981): *American sociology: theoretical and methodological structure*. University Press of America. Washington.
- Wert, J. I. (2000): "La encuesta telefónica", en García Ferrando, M.; Ibáñez, J. y Alvira Martín, F., *El análisis de la realidad social*. Alianza Universidad. Madrid: 203-217.
- Yammarino, F. J.; Skinner, S. J. y Childers, T. L. (1991): "Understanding mail survey response behavior". *Public Opinion Quarterly*, 55: 613-639.
- Yauch, C. A., y Steudel, H. J. (2003): "Complementary use of qualitative and quantitative cultural assessment methods". *Organizational Research Methods*, 6 (4): 465-481.
- Yin, R. K. (1989/2003): *Case studies research: design and methods*. Sage. Thousand Oaks, CA. (e.o. 1984).
- Zelditch, M. (1962): "Some methodological problems of field studies". *American Journal of Sociology*, 67: 566-576.

