

Serie:
Docencia Universitaria-EEES

ELABORACIÓN, ANÁLISIS E INTERPRETACIÓN DE ENCUESTAS, CUESTIONARIOS Y ESCALAS DE OPINIÓN

A. Alaminos Chica y J.L. Castejón Costa

Vicerrectorado de Calidad y Armonización Europea
Instituto de Ciencias de la Educación
Universidad de Alicante

Esta publicación no puede ser reproducida, ni totalmente ni parcialmente, ni registrada, ni transmitida por un sistema de recuperación de información, ya sea fotomecánico, electrónico, por fotocopia o cualquier otro medio, sin el permiso previo de los propietarios de copyright.

SERIE: DOCENCIA UNIVERSITARIA - EEES
Dirección de la Serie: M.^a Ángeles Martínez Ruiz

ICE/VICERRECTORADO DE CALIDAD Y ARMONIZACIÓN EUROPEA
UNIVERSIDAD DE ALICANTE

© del texto: Los autores

© de esta edición: Editorial Marfil, S.A.
C/ San Eloy, 17 • 03804 Alcoy
Tel.: 96 552 33 11 • Fax: 96 552 34 96
e-mail: editorialmarfil@editorialmarfil.com

Universidad de Alicante
Campus de Sant Vicent del Raspeig
03080 Alicante

Fotos portada: N. Sauleda Parés

I.S.B.N.: 84-268-1267-8

Depósito legal: A-495-2006

Fotomecánica, fotocomposición e impresión:
Gráficas Alcoy • San Eloy, 17 • 03804 ALCOY



ÍNDICE

PRESENTACIÓN	5
TEMA 1	
<i>Juan Luís Castejón</i>	
Planificación y Diseños de Investigación con Encuestas	7
1. <i>Delimitación conceptual</i>	7
2. <i>Métodos, diseños y técnicas de investigación</i>	8
2.1. El método experimental	9
2.2. El método cuasi-experimental	18
2.3. El método correlacional o selectivo	24
2.4. Metodología observacional: clarificación conceptual	28
3. <i>Fases generales en la realización de una investigación de encuesta</i>	30
3.1. Diseños propios de la investigación de encuesta	30
TEMA 2	
<i>Antonio Alaminos</i>	
El muestreo en la Investigación Social	41
1. <i>El todo y las partes</i>	41
1.1. La lógica del muestreo	42
1.2. Los tipos de muestreo	44
2. <i>Muestreo no probabilístico</i>	46
2.1. Muestreo de conveniencia	46
2.2. Muestreo por cuotas	47
2.3. Muestreo intencional	50
2.4. Muestreo de casos extremos	50
2.5. Muestreo de casos poco usuales	51
2.6. Muestreo de casos con máxima variación	51
2.7. Muestreo de subgrupos homogéneos	52
2.8. Muestreos estructurales	52
2.9. Muestreos con informantes estratégicos	53
2.10. Muestreo de criterio	54
2.11. Muestreo de casos críticos	54
2.12. Muestreo de casos confirmatorios y contradictorios	54
3. <i>El muestreo probabilístico</i>	55
TEMA 3	
<i>Antonio Alaminos</i>	
La dinámica de la interacción	69
1. <i>Introducción</i>	69
2. <i>La pregunta</i>	77
2.1. La formulación de las preguntas	77
2.2. La secuencia	78
2.3. Estructura	78
2.4. El entrevistado	79
2.5. Las normas sociales	79

2.6. La falsedad de la respuesta.....	80
2.7. La no respuesta.....	83
2.8. Dependencia e independencia.....	83
2.9. El cuestionario.....	84
2.10. Diseño del cuestionario.....	84
2.11. Estructura.....	85
2.12. La redacción de las preguntas.....	86
2.13. Las preguntas abiertas y cerradas.....	88
2.14. Los Formatos de cierre.....	91

TEMA 4

Juan Luís Castejón

Escalamiento: Medida de las Actitudes, Opiniones, Sentimientos y Percepciones	95
1. Introducción	95
2. Enfoques en la medición de las actitudes	96
3. Técnica Likert para la construcción de escalas de actitud.....	98
3.1. Fases en la construcción de la escala de Likert	99
3.2. Ventajas y desventajas de la técnica Likert.....	80
4. El análisis factorial en la construcción de las escalas de actitud.....	107
5. Fases en la elaboración de una escala.	
Ejemplo de construcción de una escala de evaluación de la docencia universitaria	112
5.1. Establecimiento de las dimensiones más comunes de las escalas de evaluación del desempeño docente (revisión de la literatura científica).....	112
5.2. Elaboración del cuestionario	112
5.3. Resultados del análisis factorial (método de componentes principales y rotación ortogonal-varimax).....	113
5.4. Fiabilidad de consistencia interna de los factores.....	115
5.5. Versión final de la escala.....	116

PRESENTACIÓN

La mejora de la calidad de la Universidad y la Convergencia Europea con el Espacio Europeo de la Educación Superior, llevan a la necesidad de investigar de forma continua la propia práctica educativa, una de cuyas fases es la de recabar la opinión de los estudiantes sobre múltiples aspectos de la docencia, en particular, y de los servicios universitarios en general. Uno de los múltiples ejemplos de ello lo tenemos en el documento del Consejo de Coordinación Universitaria sobre los Indicadores y Criterios de Calidad de las Nuevas Enseñanzas elaborado por la ANECA, en el que la valoración de los “resultados obtenidos a través de encuestas de opinión de los estudiantes, o procedimientos similares”, es el indicador más frecuentemente utilizado para evaluar si se cumplen los criterios establecidos.

Esta monografía tiene como objetivo general presentar los aspectos básicos del diseño, elaboración, aplicación, análisis e interpretación de cuestionarios y escalas de opinión, en el marco general de la investigación social y educativa. Como objetivos más específicos están los de conocer el proceso de diseño y elaboración de cuestionarios y escalas de opinión; definir el papel de estas técnicas de obtención de datos, dentro del proceso de investigación social y educativa; conocer y utilizar las técnicas de análisis estadístico empleadas para la elaboración de escalas y el análisis de los datos obtenidos; y aplicar los conocimientos adquiridos para la elaboración de cuestionarios, escalas de opinión, evaluación de la calidad docente, etc.

El primer tema se dedica a la planificación y diseño de investigaciones con encuestas, dentro del marco general de la investigación social y educativa, partiendo de la consideración de la encuesta como método de investigación con características propias o como instrumento de recogida de datos dentro de un

método más amplio de investigación. Una vez que en muchas ocasiones la encuesta, –entrevista, cuestionario o escala de valoración– se sitúa dentro de un proceso de investigación más amplio, como puede ser por ejemplo la comparación de las opiniones de los estudiantes sobre dos métodos de enseñanza seguidos por el profesor, nos ha parecido conveniente introducir este tema sobre el proceso general de investigación.

Un aspecto fundamental en la investigación de encuesta es el de la selección de una muestra representativa de acuerdo con los principios y procedimientos del muestreo, contenidos en el segundo tema.

El tercer tema aborda la dinámica general de la interacción en la recolección de datos, bien sea mediante la entrevista o el cuestionario, incluyendo aspectos tales como la formulación de las preguntas, la estructura de la entrevista, el diseño del cuestionario, la redacción de las preguntas, tipo, etc.

El tema cuatro introduce una graduación en la escala de medida, la elaboración de escalas de opinión que permiten obtener una valoración conjunta de un aspecto de la realidad social o educativa a partir de una serie de respuestas dadas por una persona a un conjunto de enunciados acerca del objeto de actitud de que se trate, valoración de la docencia del profesor, de los servicios administrativos prestados por la universidad, etc.

Los contenidos del tema hacen especial hincapié en los requisitos de fiabilidad y validez que deben de poseer las escalas de valoración; con especial aplicación al conocimiento de los procedimientos fundamentales de la evaluación de la docencia mediante escalas de opinión.

Aunque la estructura del libro responde a la secuencia lógica de contenidos que conforman su objeto de estudio, su lectura admite diferentes secuencias y grados de dominio en función de los conocimientos previos que posea el lector sobre los temas tratados o relacionados con el mismo. Una aproximación elemental, que no obstante permite obtener una visión coherente y significativa sobre el tema, es comenzar por la delimitación conceptual del término encuesta en cuanto procedimiento de obtención de datos, continuar con las características de las entrevistas y los cuestionarios, y abordar a continuación los conceptos básicos sobre las escalas de opinión. Otro acercamiento al tema consiste en complementar la aproximación anterior con el dominio de los procedimientos y técnicas estadísticas de análisis relativos a la elaboración e interpretación de resultados referidos a la selección de los sujetos, así como a la elaboración y análisis de las respuestas dadas a los cuestionarios y a las escalas de opinión. Finalmente, el lector puede abordar la aplicación de los conocimientos adquiridos sobre la elaboración, análisis e interpretación de encuestas, y su uso en el diseño y desarrollo de investigaciones educativas que hagan uso de estas técnicas de obtención de datos.

Los autores

TEMA 1

PLANIFICACIÓN Y DISEÑOS DE INVESTIGACIÓN CON ENCUESTAS

Juan Luis Castejón

1. DELIMITACIÓN CONCEPTUAL

Existen dos formas de conceptualizar el término encuesta:

- a) la encuesta como un *instrumento de recogida de datos* y,
- b) la encuesta *como un método de investigación*.

En el primer caso, la encuesta se considera un procedimiento más de recogida de datos, como los cuestionarios, las entrevistas, las escalas de opinión, los inventarios, etc, y forma parte de la fase de recogida de datos dentro de un método más amplio de investigación, tal como el método experimental, cuasiexperimental o correlacional. En este caso, el término encuesta es sinónimo de cuestionario o entrevista.

En el segundo caso, la encuesta se considera una estrategia de investigación, un procedimiento que implica un proceso completo de investigación, que abarca la formulación del problema, el establecimiento de los objetivos, la selección de los sujetos y el diseño y análisis de datos. En este segundo caso, el método de encuesta se entiende como una estrategia concreta dentro del método selectivo o correlacional. Según esta segunda conceptualización, se puede definir la metodología de encuestas por dos características fundamentales:

Una, como un método no experimental cuya característica principal es la recogida de información en ausencia de manipulación o intervención por parte del experimentador, utilizando procedimientos de observación o medida consistentes y estandarizados para todos los sujetos, de forma que quede garantizada la comparabilidad de los datos.

Dos, como una variante del método selectivo o correlacional en el que la finalidad de la investigación concede prioridad a los aspectos de la amplitud y representatividad de la muestra de sujetos que forman el estudio y a la generalidad de los resultados obtenidos. En la aplicación de esta estrategia cobra especial importancia el proceso de muestreo que garantiza la representatividad de la muestra y la posibilidad de generalizar los resultados obtenidos en la muestra a la población, y la obtención, por consiguiente, de una mayor validez externa de los datos y resultados obtenidos.

Desde esta perspectiva, Delanius (1988:17) define la encuesta como “*la recogida sistemática de datos de poblaciones, o de muestras de estas poblaciones, por medio de entrevistas personales, cuestionarios u otros instrumentos de recogida, especialmente cuando se refieren a grupos de personas amplios y dispersos*”

2. MÉTODOS, DISEÑOS Y TÉCNICAS DE INVESTIGACIÓN

La investigación social y educativa, en cuanto disciplinas científicas, comparten las estrategias generales del método científico, habiendo desarrollado además un conjunto de métodos y técnicas particulares adecuados a su objeto de estudio.

Los *métodos* hacen referencia a todo el proceso de investigación, adoptando cada uno de ellos una estrategia general que condiciona las posibilidades de desarrollo de cada una de sus etapas. *El diseño* se deriva del método empleado y se sitúa entre los métodos y las técnicas. El diseño de investigación establece el procedimiento concreto de obtención de datos. Mientras que *las técnicas* constituyen los procedimientos específicos para llevar a cabo diferentes etapas particulares de la aplicación de los métodos. De esta manera se habla de técnicas de documentación, técnicas de muestreo, técnicas de recogida de datos y técnicas de análisis de datos. El término metodología se refiere al estudio del método.

Los grandes métodos de investigación en ciencias sociales son el *método experimental*, el *método selectivo o correlacional* y el *método observacional*. El criterio clasificatorio se deriva directamente del grado de control interno y de la manipulabilidad o nivel de intervención que se realiza en la investigación. La intervención es máxima en el método experimental que se sitúa en un extremo, mientras que en el extremo contrario se encuentra el método observacional que presenta un mínimo control interno y un grado máximo de naturalidad. El método selectivo o correlacional se sitúa en el centro y se caracteriza por la disminución del grado de manipulación de las variables por parte del investigador y por la selección de los sujetos en función de determinadas características, como sexo, edad, nivel cultural, actitudes, etc), con la finalidad de poder estudiar la relación existente entre ellas. Como se señaló anteriormente, el método de

encuesta es una variante del método selectivo. El *método cuasi-experimental* constituye para algunos autores un método con entidad propia a medio camino entre el método experimental y el método selectivo o correlacional, mientras otros lo consideran una variante del método experimental.

2.1. EL MÉTODO EXPERIMENTAL

El método experimental tiene como finalidad establecer relaciones causales que sirven de explicación entre los hechos observados y los factores que los producen. Esta finalidad explicativa se logra a través del establecimiento de relaciones causales, lo que supone:

- a) la manipulación experimental de los hechos que se quieren explicar, creando una situación en la que se manipulan (manejan) unas condiciones antecedentes (supuestas causas) y se observan los efectos que producen en los hechos (consecuentes) que tratamos de explicar. Para establecer la existencia de una verdadera relación causal, y no una mera relación entre dos o más variables, es necesario que los cambios producidos en una variable (B) se deban a los cambios (manipulación) de otra variable (A), en la forma $A \rightarrow B$.
- b) la existencia de un control experimental de aquellos factores que puedan incidir en los resultados sin tenerlo previsto el experimentador, las denominadas variables extrañas; esta segunda característica hace que la situación experimental tenga cierto grado de artificialidad, una vez que se aíslan o controlan las variables relevantes del resto de variables que forman parte de la situación original donde se observan los hechos.

Un ejemplo de aplicación de este procedimiento lo tenemos cuando tratamos de investigar si la cantidad de ruido que hay en el ambiente incide negativamente en la atención y el rendimiento de los estudiantes, después que, en base a la observación de que los alumnos se distraen frecuentemente, se formule la hipótesis de que los ruidos de la calle que hay junto al colegio son la causa de estas distracciones.

Para ello, el experimentador, investigador o profesor dispone una situación artificial de laboratorio en la que manipula intencionalmente un aparato que produce ruido semejante al ruido ambiente, según dos condiciones de intensidad, ruido bajo y ruido alto –de igual intensidad que el de la calle–, en dos grupos diferentes, aunque semejantes en cuanto a sus características; a la vez mide el efecto del ruido sobre la atención de los alumnos aplicando una prueba estándar a todos ellos.

Para atribuir los posibles cambios en la atención al efecto del ruido únicamente y no a otras posibles causas, necesita controlar los efectos debidos a las características de los sujetos y a la situación. Como es posible que variables tales

como la cantidad de luz puedan afectar también la atención de nuestros alumnos, una forma de someter estas variables a control experimental es mantener constantes las condiciones, de forma que exista la misma cantidad de luz para todos los sujetos durante el experimento.

Campbell y Stanley (1963) identifican algunas variables extrañas potenciales, enumerando las siguientes “amenazas” a la validez interna de los resultados:

1. Historia. Se refiere a todos aquellos factores externos a los sujetos que ocurren durante el experimento y pueden afectar los resultados sin tenerlo previsto el experimentador.
2. Maduración. Se refiere a los cambios internos que se producen en los sujetos (cansancio, práctica, etc) y que afectan los resultados, haciendo que se atribuyan indebidamente los cambios observados en la variable dependiente a la influencia de la independiente.
3. Aplicación de pruebas antes del experimento, lo cual puede sensibilizar a los sujetos haciendo que no se obtengan los mismos resultados en las pruebas que se realizan para la medida de la variable dependiente después del experimento, a los que se obtendrían si no se hubieran aplicado las primeras pruebas.
4. Regresión estadística. Es un fenómeno que consiste en la tendencia que tienen los valores extremos, muy altos o muy bajos, a “regresar” hacia los valores medios en una segunda medida. Es posible por tanto que el cambio entre una y otra medida se deba a este fenómeno más que a la variable independiente, haciendo sin embargo que atribuyamos el cambio a ésta, de forma inapropiada.
5. Mortalidad experimental. Se produce cuando los sujetos abandonan el experimento por unas causas determinadas que afectan de forma diferente a uno y otro grupo.
6. Selección diferencial. Es una de las amenazas mas graves. Se produce cuando participan en el experimento dos grupos de sujetos “seleccionados”; esto es, no-equivalente o con características diferentes, lo cual no hace comparables los resultados.
7. Efecto de interacción. Se produce cuando se combinan los efectos de las anteriores variables.
8. Efecto del experimentador. Formulado por Rosenthal (1967); se produce cuando el experimentador, conocedor de la hipótesis, influye involuntariamente en los resultados para que éstos se adecuen a la hipótesis. Esta amenaza se controla mediante el procedimiento “a ciegas”, que se produce cuando los sujetos que participan en el experimento no conocen en qué condición experimental se encuentran; y el “doble ciego”, cuando ni sujetos ni experimentador conocen la condición o nivel de la variable independiente que se aplica en el procedimiento, que lo lleva a cabo otra persona.

9. Características de la demanda. Referenciada por Orne (1962); se produce cuando los sujetos adivinan la hipótesis y se comportan de acuerdo con los resultados esperados en base a ella. Sin embargo, no se ha demostrado la existencia de esta amenaza en la mayoría de los casos.

Por otra parte, existen una serie de procedimientos de control de las variables extrañas que tratan de garantizar la *validez interna* del experimento (Campbell y Stanley, 1963). Estas *técnicas de control experimental* son:

- a) Eliminación de las variables extrañas. Este procedimiento se puede utilizar en la medida en que se conozcan estas variables, lo que generalmente no ocurre.
- b) Mantener constantes las condiciones. Consiste en someter a todos los sujetos del experimento a condiciones semejantes (por ejemplo de luz, temperatura, etc.), a excepción de aquellas variables independientes que manipulamos. Una forma de controlar muchas de las variables de la situación en la que se realiza el experimento y que potencialmente pueden afectar los resultados, aparte de la variable independiente, es mantenerlas constantes para todos los sujetos.
- c) Balanceo. Consiste en equilibrar las condiciones que suponemos pueden influir en los resultados del experimento. Así por ejemplo, si en una clase donde vamos a realizar un experimento hay chicos y chicas y estimamos que esta variable puede afectar nuestros resultados, podemos, en el caso de tener que formar dos grupos, balancear su efecto asignando a cada grupo el mismo número de varones que de mujeres, de forma que el efecto de la variable sexo quede compensado y asegurada su no influencia en la variable dependiente.
- d) Contrabalanceo. Aunque su nombre pueda indicar otra cosa, no tiene ninguna relación con el balanceo. Es una técnica de control que se emplea para eliminar el efecto extraño que puede ocurrir como consecuencia de la aplicación de varios tratamientos sucesivos, cuyo orden de presentación podría afectar a los resultados. Se utiliza solo en los diseños intrasujeto, en los que un solo grupo pasa sucesivamente por las distintas condiciones experimentales o tratamientos. El procedimiento que se sigue en esta técnica es presentar cada condición a cada sujeto un número igual de veces en cada secuencia de presentación de forma que, al finalizar todas las secuencias, cada condición haya aparecido antes y después de todas las demás. Así, en el caso de un diseño intrasujeto en el que hay una variable con tres niveles los posibles órdenes de presentación son: ABC, ACB, BCA, BAC, CBA y CAB. El grupo total se divide en subgrupos, de manera que cada uno de los seis subgrupos se asigne a un orden de presentación. Trataremos de esta técnica al presentar los diseños intrasujeto.

- e) Sistematizar o convertir las posibles variables extrañas en variables independientes. Esto sólo se puede hacer cuando hay indicios de cuáles pueden ser las variables extrañas.
- f) Aleatorización. Esta es la técnica más poderosa y la más utilizada, sobre todo para controlar las posibles variables extrañas pertenecientes a los sujetos, como las características personales que son fuente de diferencias individuales. Consiste en asignar a los sujetos al azar a los distintos grupos, que se someten a cada condición experimental, mediante sorteo o utilización de una tabla de números aleatorios. Mediante este procedimiento se forman grupos equivalentes de sujetos, con características semejantes por término medio.
- g) Control estadístico. Se emplea cuando conocemos que existen diferencias entre los sujetos que pueden influir en los resultados del experimento, confundiendo los efectos de la variable independiente. Por ejemplo, cuando estudiamos la influencia de un método de enseñanza sobre el rendimiento y no hemos controlado previamente variables como la inteligencia o la motivación, podemos recurrir al control estadístico, normalmente el análisis de covarianza.

2.1.1. Fases en la realización de una investigación experimental

En la tabla 1 se presentan de forma esquemática los pasos fundamentales para la realización de una investigación de carácter experimental.

El primer paso en la realización de una investigación es la determinación del problema. Este paso incluye el establecimiento del problema a partir de alguna fuente, incluida la reflexión sobre la práctica. A él siguen la definición operativa, esto es el planteamiento del problema de forma que sea manejable; y el acceso a fuentes de documentación primarias que ofrecen información directa, como pueden ser los libros y artículos de revistas especializadas, así como a fuentes secundarias que nos informan sobre las fuentes primarias, tales como las revistas de resúmenes o los índices bibliográficos (Ortega y Fernández Dols, 1980). Finalmente, los centros de documentación nos ofrecen información sobre unas y otras fuentes con la ayuda de procedimientos informáticos.

La formulación de la hipótesis incluye a su vez una serie de procedimientos. Una hipótesis es “una formulación tentativa y concreta sobre el problema de manera que pueda comprobarse empíricamente. En el método experimental, la hipótesis de trabajo ha de establecer una relación causal entre variables de forma que esta relación sea la única explicación posible de los resultados que se predicen en las hipótesis” (Morales, 1981:92).

Determinación del problema	Formulación de la hipótesis	Comprobación de la hipótesis	Análisis y Discusión de Resultados
<p>Establecer el problema a partir de alguna fuente de problemas</p> <p>Definir el problema de forma precisa y operativa, así como sus objetivos concretos</p> <p>Acceder a fuentes de documentación bibliográfica primarias y secundarias</p> <p>Recurrir a bases de datos para la documentación (ERIC, PsycINFO, CSIC, etc.)</p>	<p>Hipótesis de trabajo</p> <p>Definición operacional de variables</p> <p>Tipos de variables</p> <p>Selección del diseño</p> <p>Asegurar validez interna mediante control experimental</p> <p>Asegurar validez externa. Selección de la muestra y población</p>	<p>Realización del experimento, según el diseño</p> <p>Procedimiento de recogida de datos</p> <p>Análisis estadístico de los datos obtenidos.</p> <p>Uso de técnicas estadísticas, según naturaleza de los datos y tipo de diseño</p> <p>Establecimiento de resultados</p>	<p>Análisis de los resultados de acuerdo con la/s hipótesis de trabajo</p> <p>Examen de otras hipótesis alternativas</p> <p>Discusión de los resultados de acuerdo a otros resultados y a la teoría de partida</p> <p>Preparar informe final para la publicación del trabajo</p>

Tabla 1. Pasos para la realización de una investigación experimental.

Una variable es cualquier atributo que puede tomar diversos valores, aunque sólo pueda tomar un valor en un momento determinado (Carlsmith, 1976). Las variables pueden clasificarse en distintas categorías según diferentes criterios (McGuigan, 1971). Se distingue así, por una parte, entre variables cualitativas y variables cuantitativas. Las variables cualitativas son aquellas que sólo designan una cualidad o un atributo, por ejemplo, alto o bajo, mujer u hombre. Las variables nominales son cualitativas. Cada variable está asociada a una escala de medida, nominal, ordinal, de intervalo o de razón.

Se distingue también entre variables estímulo externas al sujeto, como un método de enseñanza; variables organísmicas que pertenecen al sujeto, como el peso o la motivación; y variables de respuesta, como el número de ejercicios resueltos. Y finalmente, desde el punto de vista del experimento, las variables se dividen en variables independientes, variables dependientes y variables extrañas.

La variable independiente es la que manipula el experimentador, se establece como causa en la hipótesis y ocurre primero en el tiempo. Generalmente es una variable estímulo, aunque también puede ser organísmica e incluso de respuesta. La manipulación de la variable/s independiente/s puede producirse de dos formas, *intencional* y por *selección de valores*. En el primer caso, el experimen-

tador controla a voluntad los niveles de la variable, por ejemplo, variando el nivel de ruido mediante un generador de ruido o empleando un método de enseñanza determinado. La selección de valores tiene lugar cuando sólo podemos seleccionar niveles de una variable, como puede ser el caso de la asignación de sujetos de alto y bajo rendimiento a uno u otro grupo. Esta forma de manipulación se lleva a cabo sobre todo con variables organísmicas.

La variable dependiente es el resultado de la aplicación de la variable independiente y sus cambios han de deberse exclusivamente a la aplicación de ésta. Ha de reflejar de forma válida el efecto de la variable o variables independientes, lo que nos remite a la cuestión del control experimental.

Las variables extrañas son aquellas variables que afectan a la variable dependiente sin que su efecto haya sido tenido en cuenta por el investigador. Estas variables causan por tanto un efecto no deseado en los resultados del experimento (Campbell y Stanley, 1963). El control de las posibles variables extrañas se realiza a través del diseño y de las técnicas de control experimental. Ambos procedimientos tratan de asegurar la validez interna del experimento, mientras que la validez externa está referida a la posibilidad de generalizar los resultados del experimento a otros sujetos y situaciones.

La validez interna de un experimento es el grado en que podemos afirmar que los cambios en la variable dependiente se deben a los efectos de la manipulación de la variable/s independiente/s, y no al efecto de otras variables perturbadoras o extrañas (Campbell y Stanley, 1963).

Un tipo diferente de validez es la validez externa. Para Campbell y Stanley (1963), la validez externa se refiere al grado en que podemos generalizar los resultados obtenidos en un experimento, a otros sujetos distintos a los que han participado en él, y a otras situaciones diferentes a las que se lleva a cabo, como puede ser la situación natural donde se producen normalmente los hechos. Aunque para otros autores como Snow (1979), en la definición anterior, pueden diferenciarse dos tipos de validez, la validez de población y la validez ecológica, según nos refiramos a los sujetos o a la situación.

Para garantizar la validez externa de población debemos asegurarnos que los sujetos de la muestra particular con la que se ha realizado el experimento son semejantes a los de la población más amplia de sujetos a los que queremos generalizar nuestros resultados, lo que se consigue mediante la selección de una muestra lo suficientemente representativa de la población de referencia. En todo caso hemos de ser conscientes de que no es posible generalizar los resultados obtenidos en el experimento a sujetos de características distintas a los de la muestra, mientras que para preservar la validez ecológica se ha de tratar de realizar el experimento en una situación semejante a aquella en la que se van a generalizar o extrapolar los resultados. Se produce así cierta contraposición entre validez interna y validez externa.

La comprobación de la hipótesis se realiza de acuerdo con el diseño y se materializa en un procedimiento concreto. Como ya se ha señalado, el diseño de investigación establece el procedimiento concreto de obtención de datos, con la finalidad de asegurar la validez interna de los resultados obtenidos.

Una vez obtenidos los datos, éstos se analizan normalmente mediante una serie de técnicas estadísticas. Estas técnicas estadísticas varían según la naturaleza de los datos y el tipo de diseño empleado. La naturaleza de los datos incluye el tipo de escala de medida y su distribución, pudiéndose diferenciar desde una primera clasificación general entre técnicas paramétricas, cuando son conocidos los parámetros de la distribución (la distribución normal en la mayoría de los casos), y técnicas no paramétricas, cuando no se conoce aquella. La decisión por el empleo de una prueba paramétrica o no paramétrica va a estar en función de alguno o varios de los siguientes criterios:

- 1) El tipo de distribución muestral;
- 2) el nivel o escala de medida de los datos;
- 3) el tamaño de la muestra;
- 4) el tipo de inferencia o hipótesis que ha de ser probada. A continuación se presentan los resultados del análisis estadístico de forma sistemática y esquematizada.

Un paso importante en este proceso es la prueba de hipótesis a partir del análisis estadístico. Desde el punto de vista de la prueba de hipótesis, podemos distinguir entre la hipótesis nula H_0 , que es la hipótesis a contrastar que consideramos provisionalmente como verdadera, y que se formula en términos de la negación de nuestra hipótesis de trabajo H_1 , y la propia hipótesis de trabajo, que es la que queremos probar. Los resultados de un experimento nunca confirman la hipótesis de trabajo, sino que lo que nos permiten es rechazar o desconfirmar la hipótesis nula.

En el proceso de contraste de hipótesis estadísticas comenzamos por establecer una declaración o hipótesis de trabajo que cuenta a su vez con una hipótesis nula. Se trata entonces de decidir entre la hipótesis nula y nuestra hipótesis de trabajo, cuál tiene más probabilidad de ser verdadera. Para ello establecemos un criterio de decisión estadístico que minimice las posibilidades de error. Lo que hacemos es comparar el valor obtenido en nuestra prueba estadística de contraste con el valor necesario para considerar que podemos rechazar o no la hipótesis nula, para un nivel de significación α que nos impone la condición de que la probabilidad de rechazar la hipótesis nula H_0 sea igual o menor que α . Para rechazar H_0 a un nivel de significación dado, el valor del estadístico de contraste obtenido (z , t , F , etc) debe ser igual a o mayor que el estadístico de contraste correspondiente a ese nivel.

Al establecer el nivel de significación se está indicando el margen de error

que estamos dispuestos a aceptar al rechazar la hipótesis nula. Este nivel de significación se establece convencionalmente en 0,05 ó 0,01, y representa la probabilidad que tenemos de equivocarnos al rechazar la hipótesis nula. Cuando se emplea un nivel de significación $\alpha = 0,05$, aproximadamente el 5% de las veces estaremos equivocados cuando rechazamos la hipótesis nula y aceptemos la de trabajo.

El contraste de hipótesis y los posibles errores están también relacionados con el tipo de hipótesis de trabajo establecida, unilateral o bilateral, y con el tamaño de la muestra. Una hipótesis es no direccional o bilateral cuando no se afirma el sentido del efecto, por ejemplo, cuando decimos que “el ruido afecta la atención”, mientras que es unilateral o direccional cuando se establece en la hipótesis el sentido del efecto, por ejemplo, si consideramos que “el ruido disminuye la atención”. Las pruebas de significación estadística son bilaterales o unilaterales.

Finalmente, en el último paso del experimento, se realiza un análisis de los resultados en referencia a las hipótesis formuladas. Se examinan otras posibles hipótesis alternativas que pudieran explicar esos mismos resultados. Y por último se discuten los resultados en relación a la revisión de los hallazgos de otros trabajos, así como a la teoría en la que se enmarcan las hipótesis. En caso de estimarlo adecuado, se prepara un informe experimental para dar a conocer los resultados a través de cualquier medio de comunicación científica, normalmente, una revista del ámbito correspondiente.

2.1.2. El diseño experimental. Noción y Tipos

El diseño experimental establece cómo y en qué condiciones concretas se pone a prueba la hipótesis, de forma que se trate de asegurar la validez interna de los resultados (Bernia, 1979; Morales, 1981). Hay establecidos una serie de diseños según el problema a investigar y la/s hipótesis formuladas.

Una de las clasificaciones más citadas es la que realizan Campbell y Stanley (1963) y Cook y Campbell (1979). Estos autores diferencian tres grandes categorías de diseños, según el grado de validez interna que garantiza cada una de ellas, los *diseños pre-experimentales*, los *diseños experimentales auténticos* y los *diseños cuasi-experimentales*.

Para simplificar la presentación de cada uno de los diseños y hacerlo de forma esquemática, adoptamos la notación seguida por los autores citados. Según este formalismo notacional: X = Aplicación de la Variable Independiente; O = Observación o medida de la Variable Dependiente; A = Aleatorización; E = Equiparación; — = Grupo comparable, pero no-equivalente.

– Tipos de diseños experimentales

Los principales diseños experimentales auténticos se presentan en la tabla 2.

Todos estos diseños tienen en común el que los distintos grupos de sujetos que forman parte de cada condición experimental son equivalentes entre sí, lo que se consigue mediante la asignación al azar o a través del procedimiento de equiparación. En todos los casos, además, se trata de diseños que se utilizan para probar una hipótesis en la que hay una única variable independiente con dos o más niveles.

NOTACIÓN	DENOMINACIÓN
A O ₁ X ₁ O ₂ A O ₃ - O ₄	Diseño de dos grupos al azar con pre - test y postest
A X ₁ O ₁ A - O ₂	Diseño de dos grupos al azar con sólo Postest
A O ₁ X ₁ O ₂ A O ₃ - O ₄ A X ₁ O ₅ A - O ₆	Diseño de cuatro grupos de Solomon (Combinación de los dos anteriores, para control efecto pretest. Un mismo nivel de la VI)
E X ₁ O ₁ E - O ₂	<i>Diseño de dos grupos equiparados</i>
A O ₁ X ₁ O ₂ A O ₃ X ₂ O ₄ A O _n X _n O _n	Diseño de más de dos grupos con pre-test y postest
A X ₁ O ₁ A X ₂ O ₂ A X _n O _n	Diseño de más de dos grupos con sólo Postest

Tabla 2. *Diseños experimentales*

De acuerdo con el diseño de dos grupos al azar con pretest y postest, como su misma denominación indica, se asignan al azar a los sujetos a dos grupos, cada uno de los cuales se somete a una condición experimental o nivel de la variable independiente. El grupo que recibe el tratamiento es el grupo experimental, mientras que el grupo que no recibe tratamiento hace de grupo de control. Este grupo se encuentra en la misma situación que el grupo experimental, a excepción de la aplicación de la variable independiente. A ambos grupos se les toma una medida de la variable dependiente antes y después de la aplicación de la variable independiente.

Para examinar el efecto de la variable independiente se comparan O_2 con O_4 . Si existe diferencia estadísticamente significativa entre estas medidas –normalmente las puntuaciones medias de cada grupo–, estas diferencias se atribuyen al efecto de la variable independiente; siempre que no haya diferencias entre O_1 y O_3 , ni entre O_3 y O_4 .

La comparación entre O_1 y O_3 sirve para asegurarnos que no existen diferencias previas entre los grupos antes de comenzar el experimento, aunque la asignación al azar de los sujetos a cada grupo sea suficiente para defendernos de la amenaza de tener grupos “seleccionados” con características diferenciales.

El diseño de dos grupos al azar con sólo postest es el diseño más sencillo. El procedimiento de asignación al azar garantiza la equivalencia de los dos grupos.

El diseño de Solomon es una combinación de los dos anteriores. Se emplea para controlar el efecto pre-test; esto es, si la realización de una prueba previa afecta a los resultados posteriores.

En el diseño de dos grupos equiparados se consigue la equivalencia entre los sujetos de ambos grupos mediante el procedimiento de equiparación. Este consiste en seleccionar pares de sujetos con puntuaciones semejantes en la variable dependiente, o una variable relacionada, medida antes de aplicar la variable independiente. A continuación, cada uno de los sujetos de estos pares se asigna a cada grupo, lo que asegura la equivalencia, al menos en la variable de equiparación.

Los diseños de más de dos grupos se utilizan cuando tenemos una variable independiente con más de dos niveles o condiciones experimentales, como pueden ser tres métodos diferentes de enseñanza, o cuatro niveles distintos de ruido.

Un tipo de diseño ligeramente distinto es el diseño factorial. Este diseño se emplea cuando deseamos estudiar el efecto de dos o más variables independientes sobre una variable dependiente. El diseño factorial más sencillo es el 2×2 , en el que hay dos variables independientes con dos niveles cada una. Cada variable independiente recibe el nombre de factor; de ahí la denominación de este diseño.

2.2. EL MÉTODO CUASI-EXPERIMENTAL

Aunque posiblemente sea más adecuado hablar de diseños cuasi-experimentales cercanos al método experimental, dedicamos a este tipo de diseños un apartado diferente. El método cuasi-experimental se encuentra a medio camino entre el método experimental y correlacional. El método cuasi-experimental, y los diseños asociados, surge para examinar las intervenciones sociales que tienen su origen en el medio natural, haciendo uso predominante de los experimentos de campo. Pensemos en las campañas de tipo político, sanitario, educativo, etc, incluyendo la propia intervención del profesor en el grupo natural de clase.

Los diseños cuasi-experimentales se utilizan en las situaciones en que el experimentador no puede asignar a los sujetos al azar, por ser grupos de sujetos

ya formados con anterioridad; y cuando tiene un poder limitado sobre la manipulación de las variables independientes (Cook y Campbell, 1979; Peracchio y Cook, 1988).

Esto se produce en la mayoría de las situaciones educativas en las que no se pueden asignar los sujetos al azar, por estar formados previamente los grupos, por ejemplo, en razón de la letra inicial del apellido; estos grupos se consideran no-equivalentes aunque comparables por lo general. Además los tratamientos consisten en intervenciones educativas –como métodos de enseñanza, sistemas de evaluación, etc– que ya están prefijados, y sobre los que el experimentador tiene un margen limitado de actuación.

Aunque estos diseños suponen una pérdida de rigor experimental que puede afectar a la validez interna, tienen la ventaja de garantizar en mucha mayor medida la validez externa de tipo ecológico.

2.2.1. Tipos de diseños cuasiexperimentales

En la tabla 3 presentamos los principales tipos de diseños cuasi-experimentales agrupados en tres categorías, siguiendo la clasificación de Cook y Campbell (1979): los diseños que no permiten por lo general inferencias causales razonables o diseños pre-experimentales, los diseños con grupo de control no-equivalente y los diseños de series temporales.

Los *diseños pre-experimentales* no permiten por lo general inferencias causales razonables, por lo que es arriesgado atribuir los cambios, aunque estos se produzcan, al efecto de la variable independiente, debido a la cantidad de amenazas a las que están sometidos.

El grupo de *diseños con grupo de control no-equivalente*, aunque comparable, permite con mucha mayor garantía atribuir los cambios, cuando se producen, al efecto de las variables independientes, aunque cada uno de ellos sigue estando expuesto a diversas amenazas a la validez interna (Cook y Campbell, 1979).

DISEÑOS CUASI-EXPERIMENTALES	
– Diseños pre-experimentales	
X O	Diseño de un sólo grupo con postest
O ₁ X O ₂	Diseño de un grupo con pre y postes
$\begin{array}{c} X O \\ \hline O \end{array}$	<i>Diseño sólo postest con grupo de control no-equivalente</i>
– Diseños con grupo control no-equivalente	
$\begin{array}{c} O_1 X O_2 \\ \hline O_1 O_2 \end{array}$	Diseño con grupo de control no-equivalente, con medidas pre- y postest
O ₁ X O ₂ O ₃ \bar{X} O ₄	Diseño de un grupo con introducción y eliminación de tratamiento
O ₁ X O ₂ \bar{X} O ₃ X O ₄	Diseño de un grupo con tratamiento repetido
$\begin{array}{c} O_1 X+ O_2 \\ \hline O_1 X- O_2 \end{array}$	<i>Diseño de tratamiento invertido</i> con grupo de control no equivalente, con pretest y postest
$\begin{array}{c} O_1 O_2 \\ \hline O_3 X O_4 \end{array}$	<i>Diseño de cohortes</i> (sin división de tratamiento)
(sin notación)	Diseño de discontinuidad en la regresión
– Diseños de series temporales	
O ₁ O ₂ O ₃ X O ₄ O ₅ O ₆	Diseño simple de series temporales
$\begin{array}{c} O_1 O_2 O_3 X O_4 O_5 O_6 \\ \hline O_1 O_2 O_3 O_4 O_5 O_6 \end{array}$	Diseño de series temporales con grupo de control no-equivalente
O ₁ O ₂ X O ₃ O ₄ \bar{X} O ₅ O ₆	Diseño de series temporales con introducción y eliminación de tratamiento

Tabla 3. Diseños cuasi-experimentales, clasificación y tipos.

2.2.2. Un diseño cuasi-experimental básico: el diseño de dos grupos no equivalentes con medidas antes (pretest) y después (postest) del tratamiento

Uno de los diseños más empleados en la investigación educativa, donde no siempre se pueden asignar los sujetos al azar a los distintos grupos es el diseño de dos grupos no equivalentes con medidas antes y después del tratamiento, situado dentro del método cuasi-experimental. Este diseño se representa de forma esquemática de la siguiente manera.

$\frac{O_1 \times O_2}{O_1 - O_2}$	Diseño con grupo de control no-equivalente, con medidas pre- y postest
------------------------------------	--

Un ejemplo puede ser el siguiente: Al inicio del curso académico, se evalúa el rendimiento de dos grupos distintos de alumnos de una misma asignatura y que tienen un mismo profesor. En uno de ellos, el grupo experimental (X) se introduce un nuevo método de enseñanza. Al final de curso se evalúa mediante un procedimiento muy similar, el rendimiento de cada uno de los grupos.

Es el diseño más interpretable, el que menos amenazas tiene a la validez interna, de los diseños con grupo de control no equivalente. No obstante las amenazas de este tipo de diseño, más frecuentes que en el caso de emplear procedimientos que aseguren la equiparación de los grupos como el azar, varían dependiendo de los resultados obtenidos antes y después del tratamiento (Cook y Campbell, 1979).

Las amenazas a la validez interna dejan abierta la posibilidad de que existan otras hipótesis alternativas que sean más adecuadas, aun habiendo obtenido resultados congruentes con nuestra hipótesis de trabajo, como resultado del efecto de distintas variables extrañas. Cuanto mayor número de hipótesis alternativas podamos eliminar, más seguros podemos estar de que nuestros resultados son válidos. Ciertas pautas de resultados pueden hacer que dejemos de considerar más probables determinadas hipótesis alternativas, por lo que las amenazas a un mismo diseño pueden variar dependiendo de los resultados obtenidos.

2.2.3. Análisis estadístico

El análisis estadístico de este tipo de diseño puede realizarse mediante distintas técnicas.

1. Puede llevarse a cabo una prueba *t* de diferencia de medias para muestras independientes sobre las puntuaciones diferenciales (postest-pretest), que pueden ser puntuaciones directas o puntuaciones estandarizadas. En este

- segundo caso, primero se estandarizan, –reduciéndolas a puntuaciones z por ejemplo– y después se calculan las puntuaciones diferenciales.
2. También puede realizarse un análisis de varianza como si se tratara de un diseño factorial de dos factores, en el que las puntuaciones pretest se consideran un segundo factor. En este caso el diseño factorial sería de tipo mixto, con la variable tratamiento como variable entre sujetos y la variable pretest postest como variable intrasujetos.
 3. Alternativamente puede llevarse a cabo un análisis de covarianza, tomando las puntuaciones pretest como variable covariada. Este último método es más complejo que los anteriores y requiere que se cumplan diversos supuestos, lo que no siempre ocurre (para realizar los cálculos de este análisis de forma manual puede consultarse Castejón, 1997).
 4. En el caso de que existan más de dos grupos, la significación estadística de las diferencias entre las medias, una vez calculadas las puntuaciones diferenciales, se establece mediante la prueba F, del análisis de varianza *de un factor o en un sentido*.

En la tabla 4 tenemos unos datos típicos correspondientes a un diseño de este tipo.

Teniendo en cuenta que la media de las puntuaciones diferenciales correspondiente al grupo experimental, \bar{x}_1 es de 9,7 y la del grupo de control, \bar{x}_2 es igual a 6,7.

La prueba *t* de diferencia de medias se calcula mediante la siguiente fórmula.

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\left(\frac{\sum x^2_1 + \sum x^2_2}{(n_1 + 2) - 2} \right) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Donde $\sum x^2_1 = \sum X^2_1 - \frac{(\sum X_1)^2}{n_1}$

y, $\sum x^2_2 = \sum X^2_2 - \frac{(\sum X_2)^2}{n_2}$

En nuestro ejemplo, el valor de *t* es el siguiente,

Dado que $\sum x^2_1 = 969 - \frac{(97)^2}{10} = 28,1$

$$y, \Sigma x^2_1 = 969 - \frac{(67)^2}{10} = 36,1$$

$$t = \frac{(9,7 - 6,7)}{\sqrt{\left(\frac{28,1 + 36,1}{20 - 2}\right) \cdot \left(\frac{1}{10} + \frac{1}{10}\right)}} = \frac{3,0}{0,84} = 3,86$$

Sujeto	Grupo	Pre	Post	P. Diferencial
1	1	6	18	12
2	1	10	18	8
3	1	11	22	11
4	1	10	20	10
5	1	7	14	7
6	1	12	24	12
7	1	9	18	9
8	1	11	22	11
9	1	8	16	8
10	1	9	18	9
11	0	7	14	7
12	0	8	18	10
13	0	5	10	5
14	0	7	14	7
15	0	9	18	9
16	0	8	12	4
17	0	10	14	4
18	0	10	16	6
19	0	9	16	7
20	0	10	20	10

Tabla 4. Datos correspondientes a un diseño de dos grupos no equivalentes con pretest y postest.

El valor de t obtenido es mayor que el valor de t correspondiente a un nivel de significación $\alpha = 0,05$, y $[(10+10)-2]= 18$ grados de libertad, ($t > 2,10$) por lo que rechazamos la hipótesis nula de que no hay diferencias entre los grupos. Las puntuaciones alcanzadas por el grupo 1 (grupo experimental) son significativa-

mente mayores que las del grupo de control. Podemos concluir, por tanto, que el nuevo tratamiento afecta positivamente los resultados obtenidos.

El análisis estadístico se limita a comparar los valores de la variable dependiente O_1 y O_2 obtenidos en uno y otro grupo mediante una prueba t de Student de diferencias entre medias para grupos independientes o la prueba equivalente no-paramétrica, en el caso en el que consideremos que los datos no son susceptibles del primer análisis, debido fundamentalmente a que no se puede asumir la distribución normal de las puntuaciones o la homogeneidad de las varianzas.

2.3. EL MÉTODO CORRELACIONAL O SELECTIVO

El método correlacional se caracteriza por no hacer uso de la manipulación experimental ni de la asignación de los sujetos al azar (Cronbach, 1957, 1975; Alvira, Avia, Calvo y Morales, 1979). Como consecuencia, con este método es más difícil establecer relaciones causales explicativas. El método correlacional se limita a obtener datos, ordenarlos, clasificarlos y establecer relaciones, no causales, entre variables. Tiene por tanto un carácter predominantemente descriptivo.

Cuando se utiliza el método correlacional se estudian los fenómenos de interés en el medio natural en el que se producen, sin introducir ninguna manipulación experimental en la situación donde naturalmente tienen lugar, ni seleccionar grupos iguales de sujetos. De esta forma, aunque el diseño correlacional posea menor validez interna que el experimental, debido a la dificultad para controlar el efecto de posibles variables extrañas ligadas a la situación, tiene una validez externa y ecológica mucho mayor que el experimental. En el método correlacional no se provoca el fenómeno, sino que se observa, mide y registra, tal como ocurre realmente, lo cual garantiza la naturalidad del contexto.

Desde esta perspectiva, mientras el método correlacional tiene una función descriptiva y exploratoria, el método experimental puede utilizarse para comprobar el tipo de relaciones observadas en el primero. Ello no concede necesariamente primacía al método experimental, sino más bien complementariedad. Además, en muchos casos existen fenómenos que no pueden estudiarse con el método experimental, debido a la naturaleza de los hechos estudiados. Pensemos en las restricciones éticas que plantearía un diseño experimental que tratase de probar una hipótesis acerca de “si la ingestión continuada de grandes cantidades de alcohol en humanos provoca determinadas perturbaciones psicológicas”.

En principio, los hechos que pueden estudiarse con uno y otro método son los mismos, únicamente cambia el diseño y el procedimiento. Así por ejemplo, podemos estudiar la relación entre un método de enseñanza y el rendimiento académico al final de curso, empleando un diseño correlacional que se limite a medir simultáneamente cada variable y establecer la relación entre ellas. Alternativamente, podemos diseñar un experimento de dos grupos al azar donde se

aplique a uno de ellos la variable independiente “método de enseñanza” y se observen los cambios en la variable dependiente.

La estructura y pasos que sigue el método correlacional es semejante a la del experimental, aunque cambia el diseño y el procedimiento, así como el tipo de hipótesis, en las que no se establece una relación causal.

El método correlacional presenta una estructura interna menos coherente que el experimental, una vez que sus funciones y procedimientos cambian considerablemente, hablándose más de diferentes técnicas correlacionales que de diseños concretos (Morales, 1981).

Los diseños o técnicas correlacionales donde se relacionan dos o más variables pueden clasificarse a grandes rasgos (Cooley y Lohnes, 1971; Cuadras, 1981; Arnau, 1990), según el tipo de análisis estadístico empleado en, a) técnicas correlacionales, bivariadas o multivariadas; b) técnicas clasificatorias; c) técnicas predictivas; d) modelos causales. En la tabla 5 se presentan de forma esquemática las principales técnicas.

DISEÑO/TÉCNICA	Técnica de análisis
Técnicas correlacionales bivariadas	Asociación (datos categoriales): el estadístico χ^2 ji cuadrado. Coeficiente de correlación lineal r de Pearson
Técnicas de reducción de datos y estimación de dimensiones subyacentes	Análisis factorial
Técnicas clasificatorias	Análisis de cluster Análisis discriminante
Técnicas predictivas	Regresión múltiple Regresión logística Regresión canónica
Modelos causales	Modelos de estructuras de covarianza

Tabla 5. Diseños y técnicas correlacionales.

Los diseños correlacionales bivariados establecen la asociación o covariación entre dos variables. Por ejemplo, entre las calificaciones obtenidas en una determinada materia del bachillerato y las de la universidad, o la asociación que hay entre fumar y la probabilidad de padecer cáncer de pulmón. Las técnicas estadísticas empleadas para establecer la asociación o correlación entre dos variables

son distintas según el tipo de escala de medida utilizada, el carácter continuo o discreto de la medida y la relación, lineal o no entre las variables. Entre estas técnicas tenemos: a) el coeficiente de correlación r de Pearson para variables continuas medidas en una escala de intervalo o de razón; y b) el estadístico χ^2 –ji cuadrado–, que mide el grado de asociación entre variables categóricas (medidas en una escala nominal) de carácter discreto.

Las técnicas multivariadas se utilizan para establecer la relación entre más de dos variables. Entre estas técnicas tenemos el análisis factorial.

Las técnicas clasificatorias como el análisis de *cluster* (agrupamientos) y el análisis discriminante ofrecen la posibilidad de agrupar sujetos o variables según distintos tipos de afinidad empírica.

Mención aparte merecen los diseños correlacionales predictivos, también denominados *diseños ex post facto prospectivos* (Cohen y Cohen, 1983; Castejón y Navas, 1992), con los que se trata de predecir un criterio –como puede ser el rendimiento académico– en función de diferentes variables predictoras relacionadas con éste, medidas en un momento anterior. Las técnicas estadísticas más empleadas en estos diseños son las ecuaciones de regresión múltiple. La regresión logística se emplea cuando las variables dependientes o independientes son de tipo categorial, como el género; la regresión canónica se utiliza cuando hay más de una variable dependiente, como el rendimiento y la satisfacción del estudiante.

Finalmente están los modelos causales que hacen uso de las técnicas de estructuras de covarianza (Jöreskog, 1977; Kenny, 1979). Con este procedimiento se trata de inferir relaciones causales a partir de diseños correlacionales. Estos modelos han experimentado un auge reciente en el ámbito psicoeducativo (Castejón y Vera, 1996; Sternberg, Castejón, Prieto, Hautämaki y Grigorenko, 2001).

La técnica en la que se asienta el análisis estadístico de la mayor parte de estos diseños es el *coeficiente de correlación lineal r de Pearson*, empleado cuando las dos variables son continuas y están medidas en una escala de intervalo o de razón, e incluso cuando las variables están medidas en una escala ordinal, si bien en este último caso es más adecuado el coeficiente de correlación por rangos r_{rho} .

El coeficiente de correlación formulado por Pearson consiste en un procedimiento para medir el grado de covariación entre dos variables. Este tipo de coeficiente se emplea cuando las variables están medidas en una escala de intervalo o de razón, y cuando además existe una relación lineal entre las variables. La suposición de que existe una relación lineal entre las variables, –frente a una relación de otro tipo como la curvilínea, cuadrática, exponencial, etc.– es una justificación importante para el uso de este coeficiente. La linealidad de la relación entre dos variables se establece observando el diagrama de dispersión de las

puntuaciones representado en un sistema de coordenadas cartesianas en las que situamos los valores de una variable en el eje de ordenadas y los valores de la otra en el eje de abscisas.

De esta forma, si la relación es lineal, la r de Pearson es una medida de hasta qué punto los mismos individuos o sucesos ocupan la misma posición relativa respecto a dos variables. Es uno de los coeficientes de correlación más utilizados ya que la mayoría de las variables alcanzan un nivel de medida de intervalo y no está desaconsejado tampoco para establecer la correlación entre dos variables cuando están en una escala de medida ordinal.

La varianza común entre dos variables se establece en términos porcentuales como el coeficiente de determinación; esto es, como el cuadrado de la correlación de Pearson entre las dos variables. Así, un coeficiente de correlación $r_{xy} = .70$, indica que el porcentaje de varianza que tienen en común las dos variables es $r^2 = .49$. Por lo tanto, hay un 49% de varianza común entre las dos variables.

El cálculo del coeficiente r se puede llevar a cabo de diferentes formas, dependiendo del tipo de puntuaciones sobre el que se establece.

Cuando se tienen puntuaciones de desviación, esto es puntuaciones que indican la diferencia entre la media y una puntuación dada, la fórmula para el cálculo es la siguiente:

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \cdot \sum y^2}}$$

donde $\sum xy$ es el sumatorio de los productos cruzados de las puntuaciones de desviación.

Cuando lo que se tiene son puntuaciones originales o puntuaciones directas, la fórmula para el cálculo es la siguiente:

$$r = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \cdot \left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

Para ejemplificar el cálculo del coeficiente de correlación lineal podemos tomar como datos los que se ofrecen en la tabla 6.

Tomando como referencia estos datos, calculamos fácilmente el coeficiente de correlación:

$$r = \frac{775 - \frac{49 \times 91}{7}}{\sqrt{455 - \frac{49^2}{7} \times 1435 - \frac{91^2}{7}}} = \frac{138}{\sqrt{112 \times 252}} = \frac{138}{168} = 0,82$$

Como vemos la relación existente entre las dos variables es muy alta y de tipo positivo. Esto es, por lo general, para cada sujeto, cuanto mayor puntuación obtiene en una variable, mayor puntuación alcanza en la otra.

Sujeto	X	X ²	Y	Y ²	XY
1	1	1	7	49	7
2	3	9	4	16	12
3	5	25	13	169	65
4	7	49	16	255	112
5	9	81	10	100	90
6	11	121	22	484	242
7	13	169	19	361	247
	ΣX=49	ΣX ² =455	ΣY=91	ΣY ² = 1435	ΣXY=775

Tabla 6. Datos hipotéticos para ejemplificar el cálculo del coeficiente de correlación de Pearson.

2.4. METODOLOGÍA OBSERVACIONAL: CLARIFICACIÓN CONCEPTUAL

El significado del término “observacional” no es unívoco. Este término es tan amplio que puede estar referido, desde las primeras fases de obtención de datos del método experimental hasta los procedimientos de investigación cualitativa e interpretativa asentados en supuestos fenomenológicos, presentes en los paradigmas alternativos al método científico.

Dentro de la metodología científica cabe hacer otra distinción importante entre la observación como método, tal como venimos comentando, y la observación como técnica. En este sentido la observación tiene una función más limi-

tada, restringida a un procedimiento de recogida de datos que puede formar parte de cualquier método, incluido el método experimental (Anguera, 1978; 1991a).

El método observacional y el método correlacional son denominaciones equivalentes (Morales, 1981), presentando el método observacional los mismos rasgos que el correlacional; y ambos con características diferenciales respecto a la experimentación. Quizá el matiz que presenta la denominación *correlacional*, en la línea de Cronbach, está referido a la utilización predominante de escalas de medida que permitan un análisis estadístico de los datos a un nivel mayor que el nominal o categorial. La denominación observacional abarca todos aquellos procedimientos que no hacen uso de la manipulación experimental; además que, en la fase de obtención de datos, aunque no excluye la medida y el empleo de técnicas correlaciones, concede preponderancia a la observación más o menos sistemática.

Tampoco se puede establecer una separación radical entre el método experimental, como aquél que permite el establecimiento de formulaciones explicativas, y el método observacional, como método que tiene una función meramente descriptiva (Anguera, 1991a,b; Riba, 1991). Con el método observacional también se considera posible formular y comprobar hipótesis explicativas (Anguera, 1981). No hay una dicotomía observacional/experimental, sino una gradación a lo largo de una serie de dimensiones comunes a ambos métodos.

Los datos obtenidos con la metodología observacional están medidos sobre escalas ordinales y categoriales, en su mayor parte; lo que no obstante, sigue haciendo posible la aplicación de técnicas estadísticas (Anguera y Blanco, 1984). Como señalan Bakerman y Gottman (1986), Anguera (1991a) el sistema categorial puede considerarse como el instrumento de medida en la investigación observacional.

Respecto al control realizado en el método observacional, conviene señalar brevemente dos aspectos. Por una parte es posible un control moderado de la muestra, seleccionando para la observación determinados sujetos, situaciones, momentos, etc. (Feldman y Hass, 1970; Coll, 1981); por otra, es posible llevar a cabo “tratamientos naturales” donde la variable de tratamiento no está bajo el control de los sujetos, sino que ocurre de forma natural y sin intervención experimental.

En todo caso, sí hay acuerdo en considerar la importancia de la metodología observacional en la investigación psicoeducativa (Coll, 1981), debido a dos razones: a) la necesidad de estudiar el comportamiento en el contexto natural donde tiene lugar, y b) el carácter de muchas situaciones educativas en las que es difícil analizar los fenómenos que se producen, con otra metodología diferente; muchas situaciones de conducta no verbal, y sobre todo de interacción entre personas, como el juego y la interacción madre-hijo o profesor-alumno, pertenecen a este tipo.

3. FASES GENERALES EN LA REALIZACIÓN DE UNA INVESTIGACIÓN DE ENCUESTA

En la realización de una investigación con el método de encuesta se suelen distinguir cinco grandes *fases de trabajo*, que se corresponden con los grandes pasos del método científico en general:

1. Formular con precisión los objetivos de la investigación, establecer hipótesis, delimitar las variables, operativizar conceptos, y determinar y elegir la muestra mediante el procedimiento adecuado de muestreo.
2. Elección y/o construcción del o de los instrumentos de recogida de datos (información).
3. Preparación del trabajo de campo y realización de las entrevistas o aplicación de los cuestionarios.
4. Preparación de los datos (codificación, grabación, etc), análisis y establecimiento de resultados
5. Análisis e interpretación de los resultados y elaboración del informe final.

Un aspecto fundamental en la investigación de encuesta es la *selección de una muestra amplia y representativa* de la población, con el objetivo de generalizar los resultados de la encuesta (descripción de las preferencias por un producto, tendencia de voto, etc.) a una determinada población.

Asimismo, se debe de elegir *el diseño* que se va a emplear en la investigación. Este diseño está determinado por distintos factores, como son la profundidad de análisis que se desea llevar a cabo: meramente descriptivo o explicativo/predictivo; la secuencia temporal de recogida (o disponibilidad) de datos: longitudinal o transversal, etc.

Una vez que el método de encuestas constituye una variante del método correlacional o selectivo, lo que no conlleva la manipulación de variables, ni un control experimental exhaustivo, los diseños son generalmente de tipo descriptivo. No obstante, también se tratan de establecer relaciones causales en este tipo de investigación.

3.1. DISEÑOS PROPIOS DE LA INVESTIGACIÓN DE ENCUESTA

La elección de un determinado tipo de diseño viene determinada por el tipo de problema que se quiere investigar y por la hipótesis específica de la que se parte. Los diseños propios de la investigación de encuesta tenían en un principio una finalidad descriptiva y empleaban grandes muestras representativas, sobre todo en el campo sociológico.

En el ámbito educativo, han ido adquiriendo una finalidad cada vez más analítica, pasando a utilizarse para estudiar las relaciones y asociaciones entre variables así como para establecer comparaciones entre subgrupos de una misma población. De esta forma, muchos de los diseños propios del método correlacio-

nal y cuasi-experimental también se emplean en la investigación de encuesta. Incluso la combinación de una secuencia temporal de recogida de datos, junto con el empleo de poderosas técnicas de análisis de datos, como las técnicas de análisis de estructuras de covarianza (path análisis, modelos estructurales, etc), ha dado lugar al establecimiento de modelos causales a partir de datos recogidos según esquemas temporales propios de diseños correlacionales.

3.1.1. Diseños transversales

La dimensión temporal en la que se obtienen los datos de la encuesta determina la naturaleza general del diseño, pudiéndose distinguir, en base a esta dimensión, entre diseños transversales y longitudinales.

El diseño transversal se emplea para describir una población en un momento dado. En un diseño transversal se plantea, como mínimo, establecer diferencias entre distintos subgrupos de la población, así como encontrar relaciones entre las variables más importantes. Por ejemplo, ¿hay relación entre el género y la preferencia por un método de enseñanza u otro? o ¿tienen preferencia los alumnos de diferentes cursos por las clases teóricas o prácticas?.

Las técnicas de análisis de datos utilizadas en este tipo de diseño son predominantemente descriptivas. Estas técnicas incluyen índices de tendencia central (media, mediana y moda), de dispersión (varianza, desviación estándar) y de forma de la distribución (simetría, cúrtosis, etc.) de las respuestas a cada una de las preguntas del cuestionario utilizado en la investigación.

Cuando el análisis de datos combina y relaciona la información de dos variables categóricas en una tabla de contingencia, para establecer la asociación existente entre ellas, se emplea la técnica estadística χ^2 –ji cuadrado–.

En este caso lo que tenemos son frecuencias que se tabulan en una tabla de contingencia, y se trata de establecer si existe independencia o asociación significativa entre las variables. Por ejemplo, si existen diferencias entre hombres y mujeres respecto a qué es lo que les hace preferir uno u otro automóvil, la apariencia o el rendimiento.

Para ello podemos observar los datos de la tabla 7 donde se muestran las respuestas de un grupo de hombres y mujeres a la cuestión de si a la hora de comprar un coche tienen en cuenta la apariencia o el rendimiento.

El cálculo del estadístico ji-cuadrado se realiza mediante la fórmula:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Esto es, se suman a través de filas y columnas el cociente entre las diferencias de las frecuencias esperadas y observadas al cuadrado, y las frecuencias esperadas.

El cálculo de las frecuencias esperadas correspondientes a cada celdilla se

divide el valor marginal de la fila correspondiente entre la frecuencia total y se multiplica el valor obtenido por el valor marginal de la columna correspondiente. Así, los valores marginales obtenidos en nuestro ejemplo son los que se muestran en la tabla entre paréntesis.

Sexo	Apariencia	Rendimiento	Marginal fila
Hombres	75 (100)	125 (100)	200
Mujeres	150 (125)	100 (125)	250
Marginal columna	225	225	45

Tabla 7. Datos para el cálculo de χ^2 (Ji-cuadrado)

A partir de los datos de nuestro ejemplo, calculamos el valor de ji-cuadrado según esta fórmula.

$$\chi^2 = \frac{(75-100)^2}{100} + \frac{(125-100)^2}{100} + \frac{(150-125)^2}{100} + \frac{(100-125)^2}{100} = 6,25 + 6,25 + 5 + 5 = 22,5$$

Una vez que el valor de 22,5, con $(2-1)(2-1)=1$ grado de libertad, que se ha obtenido es mayor que el valor de 2,71 requerido para ser significativo al nivel de $\alpha=.05$, rechazamos la hipótesis nula. De otra manera, concluimos que los hombres y las mujeres se basan en aspectos distintos a la hora de preferir uno u otro automóvil. Y de manera más específica, que las mujeres establecen su preferencia por los automóviles en base a la apariencia, de forma más frecuente que los hombres.

Este procedimiento puede extenderse al análisis de variables donde hay más de dos niveles, pudiendo tener tablas de contingencia de 2×3 , 3×4 , etc.

La utilización del estadístico de contraste ji-cuadrado está limitada por dos factores. El primero, es que cada observación o frecuencia es independiente de toda otra observación, por lo que no se deben hacer varias observaciones en el mismo individuo y tratar cada una de ellas como si fueran independiente de todas las otras observaciones. En segundo lugar, cuando el número de frecuencias es pequeño o cuando la proporción esperada en cualquier celdilla es pequeña, pueden cometerse errores. Para que esto no ocurra, la frecuencia esperada en todas las celdillas debe ser igual o mayor que 5 cuando hay un grado de libertad, e igual o mayor que 5 el 80% de las veces cuando hay más de un grado de libertad. No obstante, si no se cumple este requisito, podemos utilizar alguna otra prueba estadística como el test de probabilidad exacta de Fisher (Siegel, 1956).

3.1.2. Diseños longitudinales

Cuando el objetivo es el de establecer el cambio que se produce en determinados aspectos en una población se utiliza un tipo de diseño longitudinal en el que se toman medidas a lo largo de un periodo de tiempo. Existen distintas variantes de este diseño.

A) *Diseño de muestreos sucesivos con grupos de sujetos distintos o diseño de encuestas longitudinales de población.*

En este diseño se realizan sucesivos estudios transversales con el objetivo de evaluar el cambio que se produce en una población en puntos sucesivos del tiempo. Los sujetos, aunque pertenecen a la misma población, no tienen por qué ser los mismos. Se trata por tanto de repetir la misma encuesta en sujetos de características muy similares. Así por ejemplo, si en la primera encuesta participan sujetos estudiantes de 18 años de ambos sexos, en la segunda deben participar sujetos con características muy similares. Asimismo, pueden encuestarse a los mismos sujetos en dos o más momentos distintos. En ambos casos se habla de *diseño longitudinal de tendencias*.

Cuando los diseños longitudinales estudian el cambio en una población *tras la implantación de algún programa*, reciben el nombre de *diseños de series temporales interrumpidas*. Estos diseños se emplean mucho en la evaluación de programas de salud, servicios sociales, educación, etc.

El análisis de datos de estos diseños se realiza mediante técnicas estadísticas específicas de series temporales destinadas a establecer el cambio de nivel o tendencia de las series, que tienen en cuenta la existencia de autocorrelaciones entre las variables, como los modelos ARIMA. Un tipo de análisis muy básico consiste en comparar la puntuación mediana de cada serie y ver si hay una diferencia notable entre las serie de observaciones en las que se ha dado tratamiento y en las que no.

B) *Diseño de cohorte*

Este diseño recibe su nombre tan marcial de las cohortes romanas. Su objetivo es el de estudiar el cambio producido en grupos de individuos que tienen una característica común muy notoria. La característica más frecuente es la edad, pero también pueden ser otras características como pertenecer al mismo curso, a la generación del baby-boom, etc.

Las cohortes son grupos de sujetos que pasan sucesivamente por las mismas condiciones ambientales a través de instituciones formales como la escuela, o informales, como la familia (Cook y Campbell, 1979); ejemplos de cohortes son los hermanos mayores respecto a los hermanos menores, o los alumnos de un curso de este año y los alumnos del mismo curso del año anterior. El diseño de cohortes puede considerarse un caso especial del diseño de dos grupos no equivalentes.

En la tabla 8 aparece la notación de las distintas variedades de este diseño con las amenazas a la validez interna de cada una de ellas.

Notación	Amenazas a la validez
O_1 $X O_2$ Diseño sencillo o básico	– Selección – Historia – Efecto de la prueba
$O_1 O_2$ $O_3 X O_4$ Diseño de cohortes (sin división de tratamiento) Diseño de ciclos institucionales	– Historia – Efecto de las pruebas – Selección
O_1 $X_1 O_2$ _____ O_1 $X_2 O_2$ Diseño de cohortes con división de tratamiento	– Selección

Tabla 8. Representación esquemática del diseño de cohortes.

El diseño sencillo compara las observaciones obtenidas en una cohorte anterior que no ha recibido el tratamiento (por ejemplo, hermanos mayores que no vieron el programa “Barrio Sésamo”, con una cohorte observada en un momento posterior que sí lo ha recibido (hermanos menores que si lo ven normalmente).

En el diseño de cohortes sin división de tratamiento se toman además medidas antes y después del tratamiento.

Veamos un ejemplo de diseño de cohortes básico. En la Comunidad Valenciana se quiere analizar el programa de educación bilingüe implantado. En concreto, se quiere investigar la influencia de la educación bilingüe en las habilidades lingüísticas de los alumnos. Para ello, se compara el rendimiento en una prueba lingüística de la primera promoción de alumnos, que ha recibido toda su formación en bilingüe, con el que obtuvieron los alumnos de una promoción inmediatamente anterior que habían cursado sus estudios según la programación previa a la entrada en vigor de la ley de Uso y Enseñanza del Valenciano.

El diseño de cohortes con división de tratamiento es una combinación de dos diseños de cohortes sencillos, en donde se establecen distintos niveles del tratamiento.

Un ejemplo de este último diseño es el siguiente. Para averiguar si el programa “Bárrio Sésamo” ejerce algún efecto sobre el desarrollo intelectual de los niños que lo ven, se recogen datos del mismo al finalizar la etapa preescolar de un amplio grupo de niños que ven el programa. Entonces se divide a los niños en edad preescolar en dos grupos, los que ven el programa entero todos los días (alta exposición) y los que sólo lo ven ocasionalmente (baja exposición). Después se establece la comparación del desarrollo de estos niños con los hermanos mayores que no veían el programa y que han sido divididos en dos grupos, de los que se supone que habrían estado en el grupo de alta y baja exposición.

Los datos de este tipo de diseños pueden analizarse mediante una prueba *t* de *diferencia de medias para grupos independientes*, un análisis de varianza para medidas independientes, un diseño factorial para el diseño de cohorte con división del tratamiento, o bien mediante la correspondiente prueba no paramétrica, en el caso en el que no se cumplan los supuestos sobre los que se asienta la prueba *t*.

C) *Diseño de panel*

En el diseño de panel son los mismos sujetos los que se encuestan más de una vez, antes y después de una intervención programática. La atención del investigador se centra en analizar la evolución de unos mismos sujetos. Además, lo que interesa fundamentalmente en este diseño es el estudio del cambio en mayor medida que la representatividad de la población, por lo que a menudo se emplean muestras pequeñas. En estos diseños, a diferencia del diseño de grupos con sujetos distintos, se puede establecer quiénes son los sujetos que han cambiado—qué características presentan— y quiénes los que no han cambiado. Se pueden estudiar las características de los subgrupos de sujetos que cambian. De cualquier forma, la técnica de panel se caracteriza fundamentalmente por la forma de estructurar la recogida de datos.

En este tipo de diseño correlacional se puede establecer si una de las variables que se pone en relación precede a la otra. La variable posterior en el tiempo no puede influir en la anterior. Por ejemplo, si se comprueba que hay una relación positiva entre la actitud conformista hacia la disciplina del Ejército y el hecho de haber sido ascendido, cabe preguntarse si es la actitud conformista la que ha influido en el ascenso o viceversa (el ascenso nos hace más conformistas). Para responder a esta pregunta se recogen datos sobre las actitudes hacia la disciplina y se establecen tres grupos en función del grado de conformismo de los sujetos. Tras el ascenso, se observan los resultados que aparecen en la tabla siguiente.

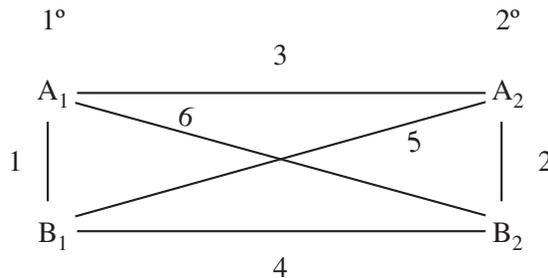
Momento temporal 1		Momento temporal 2
Conformismo		% de ascendidos
Grupo alto	20%	40
Grupo medio	44%	37
Grupo bajo	35%	22
N= 318		N= 76

Tabla 9. *Relación entre conformismo y ascenso.*

Los datos indican que son los sujetos que manifiestan mayor conformismo los que tienen mayores probabilidades de obtener un ascenso.

Un tipo especial de diseño de panel es el de correlaciones cruzadas y diferidas en el tiempo. Cuando se dispone de datos sobre la correlación entre dos variables en momentos distintos, se puede inferir el sentido de la causalidad.

El conjunto de correlaciones de dos paneles es:

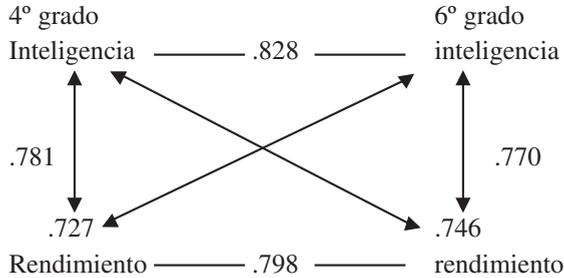


Las correlaciones 1 y 2 forman los paneles; las correlaciones 3 y 4 expresan respectivamente la consistencia de cada variable; las correlaciones 5 y 6 son cruzadas y diferidas. Estas dos últimas pueden servir para inferir la dirección de la causalidad, cuando se comparan teniendo en cuenta las otras cuatro, aunque el diseño no elimina la posibilidad de que sea una tercera variable la causa común de las dos estudiadas; por otra parte, la historia, la maduración y la aplicación de test (pretest en el primer momento temporal) pueden disminuir la validez interna.

El análisis estadístico de este diseño consiste en una prueba de la significación de las diferencias entre coeficientes de correlación.

Un ejemplo aclara el empleo de este diseño. Crano, Kenny y Campbell (1972) se plantearon si la inteligencia influía en el rendimiento escolar o es el rendimiento el que lo hace sobre la inteligencia. Estos autores obtuvieron datos

de casi 5000 niños de cuarto grado sobre los resultados de un test de inteligencia y una prueba estandarizada de rendimiento, y posteriormente en sexto grado. Las correlaciones obtenidas fueron:



La diferencia $.746 - .727$ resulta estadísticamente significativa al nivel $p=0.01$, indicando un influjo mayor de la inteligencia sobre el rendimiento, que es predominante, aunque exista retroalimentación de este último sobre aquella.

D) Modelos estructurales

Estos diseños son básicamente del mismo tipo que los diseños de panel, pero más complejos, en cuanto que: a) existen más variables medidas en un mayor número de momentos temporales, y b) existen diferentes indicadores o medidas, variables observadas, de una misma variable latente o factor, lo cual hace posible controlar el error de medida. En Castejón (1996), Castejón y Vera (1996) y Sternberg, Castejón, Prieto, Hautamäki y Grigorenko (2001) pueden verse ejemplos de investigaciones que hacen uso de este tipo de metodología.

Los modelos estructurales, como ya se ha mencionado se sitúan dentro del marco general de los métodos correlacionales, aunque el procedimiento de recogida de datos y el tipo de técnica estadística empleada, *el análisis de estructura de covarianza*, pretende inferir relaciones causales a partir de procedimientos correlacionales.

BIBLIOGRAFÍA

ANDERSON, B.F. (1967). *El método científico: Estructura de un experimento*. Alcoy: Marfil.

ANGUERA, M.T. (1978). *Metodología de la observación en las ciencias humanas*. Madrid: Cátedra.

ANGUERA, M.T. (1981). La observación (I): Problemas metodológicos. En R. Fernández Ballesteros y J.A. Carroble (Eds.), *Evaluación conductual. Metodología y aplicaciones*. Madrid: Pirámide.

ANGUERA, M.T. (1983). *Manual de prácticas de observación*. México: Trillas.

- ANGUERA, M.T. (1991a). Proceso de categorización. En M.T. Anguera (Ed.), *Metodología observacional en la investigación psicológica. Vol. I*. Barcelona: PPU.
- ANGUERA, M.T. (1991b). La observación como metodología básica de la investigación en el aula. En O. Sáez (Ed.), *Prácticas escolares. Propuestas de proyectos curriculares y de investigación-acción*. Alcoy: Marfil.
- ANGUERA, M.T. y BLANCO, A. (1984). *Análisis de datos cualitativos en Psicología Cognitiva: Sistemas alternativos de cuantificación y evaluación en el marco de una estricta metodología científica*. Proyecto subvencionado por la CAICYT, Barcelona, Departamento de Psicología Experimental.
- ARNAU, J. (1978). *Psicología Experimental. Un enfoque metodológico*. México: Trillas.
- ARNAU, J. (1990). *Diseños experimentales multivariados. Alternativa analítica a la investigación psicológica y pedagógica*. Madrid: Alianza.
- ATO, M., QUIÑONES, E., ROMERO, A. y RABADAN, R. (1989). Evaluación de programas: Aspectos básicos. *Anales de Psicología*, 5(1-2), 1-12.
- BERNIA, J. (1979). *Psicología Experimental I*. Valencia: Nau Llibres.
- BAKERMAN, R. y GOTTMAN, J.M. (1989). *Observación de la interacción: Introducción al análisis secuencial*. Madrid: Morata (Original en inglés, 1986).
- BARTOLOMÉ, M. y ANGUERA, M.T. (1990). *La investigación cooperativa: vía para la innovación en la universidad*. Barcelona: PPU.
- BELTRÁN, J. (1984). Metodología y modelos. En J. Beltrán (Ed.), *Psicología educativa*. Madrid: UNED.
- BEHAR, J. (1991). Observación y análisis de la producción verbal de la conducta. En M.T. Anguera (Ed.), *Metodología observacional en la investigación psicológica. Vol. I*. Barcelona: PPU.
- BLANCO, A. y ANGUERA, M.T. (1991). Sistemas de codificación. En M.T. Anguera (Ed.), *Metodología observacional en la investigación psicológica. Vol. I*. Barcelona: PPU.
- CAMPBELL, D.T. y STANDLEY, J.C. (1963). *Experimental and quasi-experimental design for research*. Chicago: Rand-McNally. (Traducido con el título Diseños experimentales y cuasi-experimentales en la investigación social, Buenos Aires: Amorrortu, 1973).
- CARLSMITH, D. et al. (1976). *Method of research in Social Psychology*. Addison-Ewley: Reading.
- CASTEJÓN J.L. (1996). *Determinantes del rendimiento académico de los estudiantes y de los centros educativos. Modelos y factores*. Alicante: Ediciones Club Universitario
- CASTEJÓN J.L. (1997). *Introducción a los métodos y técnicas de investigación y obtención de datos en psicología y educación*. Alicante: Ediciones Club Universitario.

- CASTEJÓN, J.L. y VERA, M.I. (1996). A causal model about the individual and contextual determinants of academic achievement. *The High School Journal*, 80,(1) 21-29.
- COHEN, J. y COHEN, P. (1983). *Applied multiple regression/ correlation analysis for the behavior sciences*. Hillsdale, NJ: LEA.
- COLL, C. (1981). Algunos problemas planteados por la metodología observacional: Niveles de descripción e instrumentos de validación. *Anuario de Psicología*, 24(1), 111-131.
- COOK, T.D. (1988). Theories of program evaluation: A short history. *Evaluación Psicológica*, 4(1), 3-30.
- COOK, T.D. y CAMPBELL, D.T. (1979). *Quasi-experimentation: design and analysis issues for field settings*. Chicago: Rand McNally.
- COOK, T.D. y SHADISH, W.R. (1986). Program evaluation: the worldly science. *Annual Review of Psychology*, 37, 193-231.
- COOK, T.D. y REICHARDT, C.S. (1986). *Métodos cualitativos y cuantitativos en investigación educativa*. Madrid: Morata.
- COOLEY, W. y LOHNES, P. (1971). *Multivariate data analysis*. Nueva York: John Wiley & Sons.
- CRONBACH, L.J. (1982). *Designing evaluation of educational and social programs*. San Francisco, CA: Jossey-Bass.
- CUADRAS, C.M. (1981). *Métodos de análisis multivariante*. Barcelona: EUNIBAR (Editorial Universitaria).
- DENZIN, N.K. (Ed.), (1978). *Sociological methods. A sourcebook, 2ª ed.* Nueva York: Mc Graw Hill.
- ELLIOT, J. (1990). *La investigación-acción en educación*. Madrid: Narcea.
- EVERSTON, C.M. y GREEN, J.L. (1989). La observación como indagación y método. En M.C. Wittrock (Ed.), *La investigación en la enseñanza II*. Barcelona: Paidós/MEC (Original inglés de 1986).
- JÖRESKOG, K.G. (1977). Structural equation models in the social sciences: Specifications, estimation, and testing. En P. Krishnaiah (Ed.), *Applications of statistics*. Amsterdam: North-Holland.
- KENNY, D.A. (1979). *Correlation and causality*. Nueva York: Wiley.
- MARTÍNEZ ARIAS, R. (1981). Métodos correlacionales en psicología clínica. En F.J. Morales (Ed.), *Metodología y Teoría de la Psicología*. Madrid: UNED.
- McGUIGAN, F.J. (1971). *Psicología experimental*. México: Trillas.
- ORNE, H.T. (1962). On the social psychology of the psychological experiment with particular reference to demand characteristics and implications. *American Psychologist*, 17,776-783
- MORALES, J.F. (1981). *Metodología y Teoría de la Psicología*. Madrid: UNED.
- ORTEGA, J.E. y Fernández-Dols, J.M. (1980). *Fuentes de documentación en Psicología*. Madrid: Debate.

- PERACCHIO, L.A. y COOK, T.D. (1988). Avances en el diseño cuasi-experimental. En I. Dendaluce (Coord.), *Aspectos metodológicos de la investigación educativa*. Madrid: Narcea.
- PÉREZ SERRANO, G. (1990). *Investigación-acción. Aplicaciones al campo social y educativo*. Madrid: Dykinson.
- RIBA, C. (1991). El método observacional. Decisiones básicas y objetivos. En M.T. Anguera (Ed.), *Metodología observacional en la investigación psicológica*. Barcelona: PPU.
- STERNBERG, R.J., CASTEJÓN, J.L, PRIETO, M.D., HAUTÄMAKI, J., y GRIGORENKO, E. (2001). Confirmatory factorial analysis of the Sternberg Triarchic Abilities Test in three international samples. *European Journal of Psychological Assessment*, 17 (1), 1-16.
- SNOW, R.E. (1979). Diseños representativos y cuasi-representativos para la investigación en la enseñanza. En F. Alvira, M.D. Avia, R. Calvo, y J.F. Morales (Eds.), *Los dos métodos de las Ciencias Sociales*. Madrid: CIS.
- STENHOUSE, L. (1987). *La investigación como base de la enseñanza*. Madrid: Morata.
- WEISS, C.H. (1972). *Evaluation research*. Englewood Cliff, NJ: Prentice-Hall (Traducción castellana en Trillas, 1980).

TEMA 2

EL MUESTREO EN LA INVESTIGACIÓN SOCIAL

Antonio Alaminos

1. EL TODO Y LAS PARTES

El muestreo ha sido sin lugar a duda uno de los elementos centrales del desarrollo experimentado por la investigación empírica de la sociedad en la segunda mitad del siglo XX. Esencialmente, consiste en la posibilidad de establecer afirmaciones sobre un conjunto de elementos partiendo de la información obtenida a partir de un número limitado y menor de elementos. Esta no es una idea nueva en la actividad científica. Efectuar una afirmación general partiendo de una evidencia particular es conocida como generalización empírica y forma parte esencial del método científico. En ese sentido, todas las consideraciones que avalan desde el punto de vista de la lógica una inferencia o generalización empírica son claramente aplicables en el proceso de muestreo destinado a recoger información empírica de la realidad social. Es importante recordar este aspecto en la medida que con frecuencia el investigador se concentra en demasía sobre algunos aspectos concretos del muestreo, por ejemplo tamaño muestral o error, y descuida otros que son esenciales para el adecuado funcionamiento del método muestral: generalizar con una cierta confianza. En definitiva, inducir. Junto a la potencia de generalización, será la posibilidad de cuantificar los diferentes aspectos del diseño en términos de probabilidad (gracias a la estadística inferencial) lo que aportará la utilidad decisiva.

En primer lugar, vamos a considerar la lógica que existe detrás del muestreo y que le vincula directamente con la actividad más familiar para el investigador: inferir o generalizar. Posteriormente, distinguiremos entre los muestreos que se basan esencialmente en la generalización lógica, de aquellos otros cuya inferencia es de carácter probabilístico.

1.1 LA LÓGICA DEL MUESTREO

Como afirmamos anteriormente, mediante la inducción se pretende llegar a conclusiones que exceden la información contenida en las premisas. Un presupuesto de la ciencia es su intención de abstraer y generalizar, es decir, determinar patrones regulares que subsumen la variabilidad de las observaciones particulares. En la inducción nos interesamos, como hace la deducción, por la relación existente entre las premisas y las conclusiones. Sin embargo, analizar un razonamiento inductivo es mucho más complicado dado que debe evaluarse el grado en que las premisas soportan las conclusiones; aun cuando las premisas de un argumento inductivo correcto sean verdaderas, la conclusión solamente puede ser probablemente verdadera. Esta es precisamente la esencia del muestreo, sea probabilístico o no.

La generalización inductiva afirma algo acerca de una clase entera de objetos o sucesos, en base a información sobre una parte de la clase. Por ejemplo, sobre la base de cien observaciones, cien jóvenes, podemos inducir la distribución o intensidad de alguna característica. Por ejemplo, supongamos que todos los jóvenes (p) observados son aficionados al fútbol (q). Expresado más formalmente:

- Todos los miembros **observados** de p son q
- luego, **todos** los p son q .

Es decir

- Todos los miembros **observados** entre los jóvenes son aficionados al fútbol
- luego, **todos** los jóvenes son aficionados al fútbol

Esta parece ser una conclusión aceptable cuando todos los casos observados presentan las mismas características. Es importante apreciar que no afirmamos algo acerca de lo que conocemos, de los casos observados, sino también de los casos que desconocemos. No obstante, una situación más complicada se presenta cuando afirmamos que un porcentaje de los casos observados presentan una característica e inferimos que un porcentaje semejante de la clase total poseerá dicha característica. Por ejemplo, si el 80% de los jóvenes observados son aficionados al fútbol, inferimos que un porcentaje equivalente (el 80%) de todos los jóvenes (incluidos lo no observados, evidentemente) son aficionados al fútbol; la diferencia con el primer argumento es que en él un 100% de los casos observados poseían una determinada característica y por lo tanto ese 100% se transformaba en “todos” para el resto de la clase. En este segundo caso contamos con una fracción de ese todo, con lo que se abre la puerta a una incertidumbre. Por ello, la inducción nos permite afirmar:

- El 80% de los jóvenes **observados** son aficionados al fútbol*
- luego, probablemente el 80% de **todos** los jóvenes son aficionados al fútbol*

Como en todas las inferencias inductivas –y el muestreo es una operativización de ello– las generalizaciones inductivas van más allá de los hechos observados, considerando que no se observa toda la población. Por ello, lo mejor que se puede decir de una afirmación inductiva basada en una observación muestral es que será probablemente cierta. Nos estamos refiriendo, si bien desde el ámbito de la inducción lógica, a los mismos problemas que se consideran en el ámbito de la teoría del muestreo.

Precisamente algunos principios de la teoría del muestreo garantizan hasta cierto punto que algunos argumentos inductivos sean más sólidos que otros. En principio, la solidez de un argumento inductivo depende de lo razonable que sea suponer que los elementos observados (pertenecientes a una determinada clase) sean representativos de la clase completa. Toda aquella estrategia que garantice la similitud entre clase total y observaciones tenderá a apoyar la solidez de un argumento inductivo. Barker (1974) sugiere las siguientes reflexiones, donde las tres primeras son totalmente pertinentes para el caso del muestreo.

- a) En general, cuanto más parecidas sean las observaciones, en características diferentes a las evaluadas, más débil será la inferencia. Es decir, si los casos son homogéneos en muchas características, tenderán a parecer casos particulares dentro de una clase total más diversa.
- b) Y la afirmación complementaria, cuanto más diversos son los casos, es decir, heterogéneos en características diferentes a la considerada, más sólido será el argumento inductivo.
- c) Asimismo, en principio, cuanto mayor sea el número de observaciones que apoyan la inferencia, más sólido es el argumento, siempre que se suponga una asociación entre el incremento del número de casos y el incremento de la heterogeneidad de los casos.
- d) Otro criterio importante es que cuanto más vinculada esté la hipótesis con otras de importancia para el área de conocimiento, más sólido será el argumento. Así, si postulamos una hipótesis entre sindicalismo y partidos políticos, podría estar relacionada con otras sobre ideología política y posicionamiento social.
- e) En todo caso, cuanto más amplia o extensa es la generalización, menor será la probabilidad de ser aceptada, y su argumentación será más débil. Así, para el caso de “todos los jóvenes son ...” su probabilidad de ser aceptada es baja. Si se restringe la generalización a sólo un segmento, por ejemplo, “todos los jóvenes estudiantes en la ciudad de ...” el argumento tendrá más probabilidades de ser aceptado.

En definitiva, para evaluar la fuerza de un argumento inductivo, Barker propone que deben considerarse de modo conjunto los cinco factores mencionados. Así, hay que evaluar el número de observaciones que da pie a la inducción, en conjunción con el grado de heterogeneidad de dichos casos, el alcance y la pre-

cisión de la generalización inductiva, y todo ello en relación al conocimiento establecido con anterioridad. Estas orientaciones son totalmente pertinentes para el muestreo y especialmente en el caso de los muestreos no probabilísticos.

1.2. LOS TIPOS DE MUESTREO

Fundamentalmente, se pueden distinguir dos estrategias muestrales bien diferenciadas: el muestreo probabilístico y el no probabilístico. Ambas poseen características y cualidades que le son específicas. El muestro probabilístico es la esencia de la encuesta de opinión pública, dado que constituye la clave que permite efectuar inferencias sobre el resto de la población. La posibilidad de poder hacer estimaciones o inferencias sobre las características de la sociedad a partir de la consulta de un número muy inferior de individuos ha sido una de las claves de la investigación empírica actual.

El mayor potencial de la recogida de datos coincide, con la aplicación del muestro aleatorio a la investigación social. Este fue desarrollado por Bowley (1915) en las encuestas que efectuó entre 1912 y 1914 siguiendo un razonamiento muy simple que incluimos aquí por su interés, al contener los elementos claves del muestreo. *“There is very little doubt that the households were so chosen as to make a fair and unbiased sample of that part of the population that lives in private houses. For each town a list of all houses as given in a directory...was obtained, and without reference to anything except the accidental order... in the list, one entry in twenty was ticked. The buildings so marked, other than shops, institutions, factories, etc., formed the sample. Very strict instructions were given that no house which was occupied should be omitted, however difficult it was to get information... The adequacy of a fair sample of one in twenty house in twenty in every street in a town must give a composite result that is closely related to a result based on complete information, and that there is no reason why any percentage or average obtained should be in excess rather than in defect. It is further evident that the larger the number taken in the sample the more accurate will be the picture, and that the more general features will be presented with less uncertainty than the less common... Thus, if 800 houses are examined in a town containing 16.000 houses, and 10 per cent in the town are four-roomed, then 10 per cent of 800, i. e. 80, is the most probable number of four-roomed houses that will be found in the sample (though such exactness will not often be reached), and it is more likely that 75 or 85 will be found than 80 or 90.”*

Sobre esta idea fundadora, se desarrolla una de las estrategias de investigación más productivas para la investigación social. Esto no obvia, sin embargo, el que existan otras muchas posibilidades de diseñar muestras no probabilísticas que a su vez aportan información muy significativa al investigador. Es, sin embargo, un error grave el pretender emplear muestreos no probabilísticos como

si poseyeran las mismas propiedades que los probabilísticos y viceversa. Así, el muestreo probabilístico permite efectuar inferencias sobre la población en estudio, y es especialmente útil cuando se trata de *cuantificar* un fenómeno social, ya sea éste una opinión, una característica sociodemográfica, etc. Por su parte, el muestreo no probabilístico tiene como potencialidad la *cualificación* de los fenómenos sociales. Es un muestreo caracterizado por la intencionalidad del investigador, que intenta localizar los casos que pueden proporcionar un máximo de información. A partir de aquí, es fácil comprender la diferencia metodológica esencial entre estos dos tipos de muestreo. Son dos tipos de muestra que se complementan en lo que se refiere a obtención de la información, ya sea la intención del investigador el medir y describir o de cualificar y comprender un fenómeno social.

En un muestreo probabilístico, todos los casos presentes en el “*marco muestral*” tienen teóricamente la misma probabilidad de estar en la muestra final, o al menos una probabilidad que es conocida para cada caso. Todos los casos tienen la misma probabilidad de ser elegidos para formar parte de la muestra. En otras palabras, para el investigador es indiferente la elección de un caso u otro. Precisamente la filosofía del muestreo probabilístico es efectuar una indiferenciación con respecto a los individuos, que aparecerán aleatoriamente representados en la muestra.

Por el contrario, en un muestreo no probabilístico lo interesante es seleccionar aquellos casos que proporcionen una mayor cantidad de información, con el máximo de calidad que sea posible. El método no probabilístico se caracteriza por que no es indiferente quien forme parte de la muestra, como es el caso de los muestreos intencionados, o porque todos los casos no tienen la misma probabilidad de formar parte de la muestra, caso de los muestreos de conveniencia. Un muestreo probabilístico es ciego a la hora de seleccionar los elementos muestrales finales, mientras que la selección en un muestreo no probabilístico intencional se requiere conocer previamente las características de los individuos que forman la muestra. Esta es otra característica importante. El muestreo probabilístico no requiere un conocimiento elevado de la población que se va a estudiar, mientras que el muestreo intencional acostumbra a requerir bastante información sobre ella. Cada uno de estos tipos de muestreo es conveniente para diferentes estrategias de investigación, o en diferentes etapas de estas, y ninguno de ellos es mejor que el otro por sí mismo, sino que cada uno de ellos constituye el enfoque óptimo para un problema determinado. De hecho, nada impide que en una misma investigación coincidan ambos tipos de muestreo, efectuando una triangulación sobre algún concepto o materia en estudio.

Las técnicas y métodos de investigación son útiles para solventar problemas de investigación y no existe ninguna que sea óptima para cualquier situación. No existe la técnica de investigación “*todo terreno*”, como tampoco lo es el mues-

treo, sea este probabilístico o no. Cada situación debe de ser evaluada en la fase de diseño y resultar en tantas técnicas y métodos como requiera para la obtención de la información que se necesita, considerando siempre el equilibrio óptimo entre el coste de la investigación y la importancia de los resultados que se vayan a conseguir. Como todo investigador sabe, el coste de la investigación es el auténtico límite al diseño de ésta, imponiendo restricciones sobre lo que es posible hacer. El investigador no puede, sin embargo, en ese probable sacrificio de estrategias complementarias de investigación, someterse a la buena o mala imagen de cada técnica. El criterio debe de ser emplear la técnica que afronte óptimamente la obtención de la información que se requiere y no otro. Esta es una situación demasiado frecuente, donde el que solicita la investigación “*quiere*” que se aplique una técnica concreta, cuando para el investigador es evidente que esa información necesita, exige, otras técnicas. El conflicto no es fácil de solventar, pero la solución nunca debería de pasar por ocultar el hecho de que la técnica que se pide no es la correcta, aún cuando su realización sea más rentable.

2. MUESTREO NO PROBABILÍSTICO

En general, se pueden considerar tres formulaciones bien diferenciadas en la teoría de un muestreo no probabilístico. Se trata del muestreo de conveniencia, el muestreo por cuotas y el intencional. El muestreo de conveniencia está en los límites más periféricos del concepto metodológico de muestreo. En lo que se refiere al muestreo por cuotas posee una estructura de muestreo probabilístico siendo un sucedáneo económico de éste, mientras que el muestreo intencional es el que desarrolla auténticamente, y en un mayor grado, la potencialidad de un muestreo no probabilístico. En lo que se refiere a la realización de encuestas sociales, los más extendidos entre los no probabilísticos son el muestreo por cuotas para efectuar muestreos, y el de conveniencia para testar cuestionarios de nuevo diseño.

2.1. MUESTREO DE CONVENIENCIA

El muestreo de conveniencia es también llamado fortuito o accidental. Consiste, simplemente, en que el investigador selecciona los casos que están más disponibles. Es un muestreo fácil, rápido y barato, pero los resultados que se obtengan no pueden generalizarse más allá de los individuos que componen la muestra. Un ejemplo de esto es cuando una emisora de radio o t.v. pide la opinión sobre algo a los individuos que va seleccionando en la calle. Ese tipo de muestreo de conveniencia no permite ninguna inferencia sobre la población, y las conclusiones deben de referirse exclusivamente sobre la misma muestra. Las encuestas por internet, sin marco muestral y diseño previo, como las que efectúan los medios de comunicación son ejemplos de este tipo de muestreo. Hacer

alguna generalización a partir de dichas opiniones no posee ningún respaldo metodológico. Este tipo de muestras solo es justificable si la investigación se encuentra realmente en una fase exploratoria y se están considerando posibles desarrollos, pero dejará de serlo en el momento en que se produzca alguna generalización. Por ejemplo, es incorrecto afirmar tras un muestreo de esta clase que “*el hombre de la calle opina...*”, dado que no se posee ningún control sobre los sesgos que se produzcan. Habitualmente, los muestreos de conveniencia se utilizan para efectuar los pretest en los primeros diseños de cuestionarios, donde lo que se pretende es comprobar si las preguntas son comprensibles, el que las respuestas que se ofrecen son lo suficientemente exhaustivas, o al menos cubre la mayor parte de las posibles, si son lo bastante excluyentes, si las tarjetas son fácilmente legibles, etc. En este uso es donde los muestreos de conveniencia o accidentales son más útiles para la investigación social, si bien metodológicamente es conveniente que un cuestionario se comprueba sobre individuos con características semejantes a la que será la muestra definitiva.

2.2. MUESTREO POR CUOTAS

Muestreo por cuotas es una denominación común de una serie bastante amplia de variantes, la mayoría de las cuales se asemeja a una muestra aleatoria estratificada con afijación proporcional. Todas ellas poseen una característica común, que las diferencia del muestreo probabilístico puro. Esta estriba en que el muestreo por cuotas rompe, en algún momento, el proceso de aleatoriedad. Generalmente los muestreos por cuotas más usuales rompen la cadena de aleatoriedad en la fase de extracción del elemento muestral, es decir, en la selección del entrevistado.

Podemos considerar tres pasos básicos en el diseño de un muestreo por cuotas. Como en el caso del muestreo probabilístico estratificado, el muestreo por cuotas comienza seleccionando unas características de control, tales como edad, sexo, región geográfica, etc. Las variables de control se escogen generalmente, primero, porque se cree que están relacionados con la característica o variable a investigar, y segundo, porque existe información sobre su distribución. Se emplean, pues, dichas variables de control para estratificar la población y se determina la proporción de población que contiene cada estrato. En la muestra debe aparecer representado cada estrato; por ello, en un segundo paso, se selecciona generalmente una representación proporcional de cada estrato en la muestra final. Hasta aquí el proceso es bastante semejante al que se seguiría en un muestreo probabilístico, surgiendo la diferencia en el tercer paso, la elección del entrevistado. Lo habitual en un muestreo por cuotas es que se entregue al entrevistador la “*cuota*” de entrevistas que debe de efectuar, indicándole qué características deben de reunir los entrevistados de acuerdo a las variables de control que se emplearon para estratificar la población y afijar la muestra. De este modo,

el entrevistador queda con bastante libertad para decidir cuál es el entrevistado definitivo, aún dentro de las restricciones que le impone el perfil que marca la cuota que tiene que cubrir. Es en este momento que se rompe la cadena aleatoria, al introducirse los sesgos de selección del entrevistado. En todo caso, la muestra final reproduce en miniatura la población de acuerdo a las características que se hayan empleado como controles para la estratificación. Por ejemplo, estratificación de la población por sexo y tramos de edad. La muestra a obtener mediante cuotas debe reproducir la pirámide de población en pequeña escala.

La muestra final tendrá la misma composición que la población para las características que definen los estratos. Este método de muestreo asume que una muestra que se parece a la población, con respecto a características importantes, lo hará de igual modo con respecto a las demás que se quiere estudiar, especialmente si las características de control que definen el sistema de estratificación, y con ello de cuotas, está asociado de algún modo a la característica que se estudia. De todos modos, esta presunción es bastante arriesgada porque, y en especial en las encuestas de opinión pública, una muestra puede reproducir muy bien determinadas características de la población, pero no representarla en absoluto respecto a determinadas opiniones.

Entre las principales ventajas del muestreo por cuotas se encuentra su bajo coste, además de que puede asegurar la heterogeneidad y proporcionalidad de la muestra. Si las cuotas están definidas por variables que son importantes para el estudio, tal como raza en un estudio de racismo, el sistema de cuotas garantiza la presencia de esas dimensiones sociales, y con ello la heterogeneidad de la muestra. Sin embargo, este tipo de muestreo presenta grandes inconvenientes, especialmente en la medida que pretende cubrir los mismos aspectos que los muestreos probabilísticos, pero también en cuestiones prácticas de diseño muestral y de trabajo de campo. Lo primero es la dificultad de diseñar un sistema de estratificación para construir las cuotas que considere más de tres variables de control. La razón es bastante evidente, dado que el número de variables a considerar actúan como una función multiplicativa de cada una de las categorías dentro de la variable. Así, si empleamos sexo (dos categorías), clase social (tres categorías) y edad (cuatro categorías), obtendríamos $2 \times 3 \times 4 = 24$ tipos de entrevistados a seleccionar. La incorporación de nuevas variables de estratificación eleva sensiblemente las tipologías de entrevistados. Con ello, el número de dimensiones en las que puede ser económico el reproducir en escala la sociedad, en el diseño muestral, es bastante más limitado que en su planteamiento teórico. Otra desventaja viene dada por el hecho de que a cada entrevistador se le adjudica una cuota de personas a entrevistar, por ejemplo, tres mujeres [<dos con más de 40 años (una soltera y otra casada)> y <una con menos de 40 años (casada)>]. Tras esto, se le deja en libertad para elegir a los entrevistados siempre que reúnan estos requisitos. Este hecho incorpora bastantes errores o la posibilidad

de que se produzcan. Por ejemplo, el hecho de que el entrevistador recurra a entrevistar a sus amigos en una proporción excesiva, con lo que reproduce en la muestra sus propios gustos o afinidades, o que efectúen las entrevistas donde le sea más fácil completar sus cuotas, tal como en mercados, universidades, salida de cines, etc. Si se les impone la restricción de efectuar las entrevistas a domicilio, tenderán a elegir las casas que les parezcan más atractivas situadas en buenos barrios, o de clase media alta. El conjunto de sesgos que se puede producir es evidente. Al escoger amigos, porque en general los amigos se parecen, con lo que sesga la muestra. Si se efectúan las entrevistas en zonas que aglomere gran cantidad de gente, puede sesgar la muestra en la medida que las características de este tipo de entrevistados sea muy diferenciadas de los que no acuden a ellos. Por último, el elegir barrios que parezcan más accesibles y poco conflictivos, producirá un sesgo en el sentido de sobrerrepresentar determinadas clases sociales. En conjunto, estos sesgos en la selección de entrevistados puede influir en aquello que ya observamos: la reproducción muestral de la población a partir de determinadas características no garantiza la equiparabilidad de opiniones o actitudes.

Son muchas las estrategias empleadas para reducir los sesgos que introduce el entrevistador al seleccionar a los entrevistados. Así, por ejemplo, efectuar un control rígido sobre las áreas geográficas o barrios donde deben de efectuarse las entrevistas. Una de las opciones es efectuar un muestreo probabilístico polietápico con cluster, seleccionando los entrevistados mediante cuotas en la última etapa. Por ejemplo, se puede seleccionar una muestra de regiones dentro del estado, de provincias dentro de las regiones seleccionadas, de ciudades dentro de las provincias seleccionadas, y de barrios dentro de las ciudades seleccionadas en esas provincias. Finalmente, dentro de cada barrio, extraer una cuota controlada por edad y sexo. El resultado es una muestra no probabilística, pero con fuertes restricciones geográficas sobre la acción del entrevistador.

Sin embargo, ninguno de estos controles puede evitar el que al desconocer las probabilidades que tiene cada caso de ser seleccionado, no se puede estimar los errores debidos al muestreo. Esta incapacidad de estimar los errores muestrales es lo que diferencia a un muestreo por cuotas de otro probabilístico. Al no ser posible cuantificar los errores de tipo muestral que se incorporan a los resultados, se arroja una incertidumbre notable sobre los estimados que se realicen desde una muestra por cuotas. No obstante, todos los intentos para mejorar la fiabilidad del muestreo por cuotas se justifican en el hecho de que un muestreo probabilístico puede no ser factible, dado su elevado coste en relación a la información que se desea obtener. En esos casos, el muestreo por cuotas es una estrategia válida, en la medida que siempre es conveniente tener alguna información a carecer por completo de ella. Existen, en todo caso, muchos factores que pueden favorecer el muestreo por cuotas, tales como el hecho de que la

población sea muy homogénea con respecto a la característica que se desea medir, o que el tema de estudio pueda provocar un alto índice de no respuesta. Es el investigador el que debe de decidir, ponderando todos los factores, si la opción de un muestreo por cuotas es válida o no para el tipo de información que se desea recoger.

2.3. MUESTREO INTENCIONAL

El muestreo intencional constituye una estrategia no probabilística válida para la recolección de datos, en especial para muestras pequeñas y muy específicas. Como afirma Quinn (1988) *“La potencia del muestreo estadístico depende de seleccionar una muestra verdaderamente aleatoria y representativa que permita hacer generalizaciones desde la muestra a una población mayor. La potencia en el muestreo intencional está en seleccionar casos ricos en información para estudiar en profundidad”*. Existen varias estrategias de muestreo intencional y todas ellas son procedimientos para alcanzar casos que poseen mucha información cualificada sobre el tema que se estudie. Por otra parte, los muestreos intencionales son especialmente útiles en la caracterización de formas sociales y muy débiles en la cuantificación de la distribución de una variable. Precisamente, lo contrario sucede con los muestreos probabilísticos. En la medida que se pretenda caracterizar formas sociales, un buen indicador de la calidad de la información y de que se puede dejar ya de recoger datos, es cuando la información que se obtiene es repetitiva y no aporta nada nuevo, o lo que aporta es muy poco relevante. Por ejemplo, en el estudio de Cuber y Haroff (1968) sobre estilos de vida de la clase media norteamericana, la recolección de datos se dio por terminada cuando comenzaron a recoger información redundante. Dado que su interés se centraba en las formas que adoptaban los estilos de vida y no en su distribución, la redundancia de la información no aportaba nada nuevo para cualificar las tipologías y dieron por terminada la recolección de datos. Al mismo tiempo, esa redundancia aportaba validez a la información que se había recogido.

Las estrategias de muestreo intencional son muy variadas, respondiendo en su mayoría a problemas y enfoques de investigación muy concretos. Vamos a considerar seguidamente algunas de las empleadas más frecuentemente así como alguna de sus variantes.

2.4. MUESTREO DE CASOS EXTREMOS

El muestreo de casos extremos consiste en seleccionar solo aquellos casos cuyos valores se encuentran en el extremo del rango de una variable. Por ejemplo, si la variable es ideología política expresada en términos de 10 posiciones (desde 1 en la extrema izquierda, hasta 10 en la extrema derecha), un muestreo de casos extremos estaría dirigido a recoger los individuos que se autoubican ideológicamente en el 1 y en el 10. En este ejemplo concreto, una investigación

efectuada por el autor¹ aplicando un diseño muestral intencional que integraba diferentes estrategias, los resultados obtenidos fueron bastante satisfactorios al cualificar los perfiles de estos casos extremos. Los problemas que se encontraron fueron los habituales en este tipo de muestreos, donde la interacción “*cara a cara*” con casos extremos recoge muchas dificultades. En este caso, la misma técnica de entrevistas ‘*focalizadas*’ (*focused interview*) abiertas facilitó tanto el hecho mismo de la recolección como la interpretación posterior.

Este tipo de muestreo es especialmente útil cuando el investigador se encuentra interesado en descubrir las fronteras de la acción social o de las instituciones. Muchas veces se puede aprender más del estudio intensivo de casos extremos que de descripciones estadísticas de casos promedios. Precisamente, la aplicación de los métodos y técnicas a países en vías de desarrollo es, en algún aspecto, un ejemplo de lo mucho que se puede aprender del estudio de casos extremos.

2.5. MUESTREO DE CASOS POCO USUALES

El muestreo de casos poco usuales o raros consiste en seleccionar aquellos elementos cuyos valores en el rango de una variable son poco frecuentes. Existen muchos tipos de variable social donde la aparición de alguno de sus valores es poco usual, por lo que sería poco eficiente en tiempo y coste emplear cualquier muestreo diferente, ya sea probabilístico o no. Tal como lo ejemplifica Smith (1975), para poder estudiar el impacto de las vasectomías en la vida familiar, sería absurdo cualquier estrategia que no consista en localizar expresamente casos con esa característica para formar la muestra. Por ejemplo, ir a los archivos de una o varias clínicas que efectúen vasectomías y extraer de allí la muestra.

Es evidente que difícilmente se puede generalizar al resto de la población, y que el emplear datos de clínicas puede sobrerrepresentar determinados estratos sociales, pero también lo es el que un muestreo probabilístico sobre este tema corre el riesgo (y más aún en 1975) de no recoger ningún caso.

La lógica que subyace a este tipo de muestras es evidente: Cuando los casos que pueden encontrarse en relación a un tema de estudio son poco frecuentes, la necesidad de un muestreo específico e intencionado de estos casos es un argumento que se sustenta por sí solo.

2.6. MUESTREO DE CASOS CON MÁXIMA VARIACIÓN

El muestreo de casos con máxima variación pretende construir una muestra lo más heterogénea posible. Esto que en muestras pequeñas representa un pro-

1. A. Alaminos “Análisis de consistencia de las escalas de autoubicación ideológica en España”. Informe C.I.S. (1987).

blema, se convierte mediante este tipo de muestreo intencional en una potencialidad, partiendo del siguiente principio: Cualquier pauta común que surja de una gran variación posee un interés particular. Así pues, el propósito central en un muestreo de máxima variación será, paradójicamente, buscar la homogeneidad que puede suponer la existencia de pautas comunes.

El procedimiento para conseguir la máxima variación en la muestra consiste en seleccionar una serie de características que serán empleadas para estructurar el diseño de esa diversidad que se pretende, por ejemplo, las variables raza, religión, cultura, etc. Esas variables son las que van a determinar la composición de la muestra, y con ello garantizar, al igual que en el sistema de cuotas, la presencia de esa diversidad de características.

Una vez que se ha diseñado la muestra a partir de dichas características, es necesario emplear un procedimiento de recolección de datos que implique una descripción detallada y exhaustiva de cada caso en relación al tema objeto de estudio. Por ejemplo, de aplicar entrevistas en profundidad, es aconsejable emplear una guía de la entrevista lo bastante detallada, que estructure lógicamente la recogida de la información y potencie la comparabilidad entre los casos.

Una vez en el análisis, la idea central es concentrarse en la detección de pautas comunes compartidas entre los elementos de la muestra. La fiabilidad, validez e importancia de los hallazgos que se obtengan vendrán avalados, precisamente, por la heterogeneidad intencionada de la muestra.

2.7. MUESTREO DE SUBGRUPOS HOMOGÉNEOS

Una estrategia de muestreo intencionado que emplea una lógica semejante, si bien en sentido inverso, es el de escoger una muestra pequeña lo más homogénea posible. En el desarrollo de una investigación puede surgir la necesidad de recoger información en profundidad sobre uno o varios grupos. En este tipo de muestra se reúnen una serie de individuos de condiciones sociales y experiencias semejantes. Obviamente, la saturación de la información que se quiere conseguir se obtiene relativamente más rápido en el seno de un grupo homogéneo, que efectuando la recolección de datos a los individuos aislados que componen la muestra. Así, por ejemplo, una técnica que se apoya sobre este tipo de muestreo intencional es la entrevista colectiva, también llamada grupo de discusión. Este tipo de técnica se caracteriza por la recolección en conjunto de la información generada desde la muestra. La composición de ésta se define por criterios técnicos internos a la investigación o mediante estratificación de algún tipo, como área geográfica, sexo, edad, etc.

2.8. MUESTREOS ESTRUCTURALES

Los muestreos estructurales se caracterizan, a diferencia de los anteriores, porque los individuos que componen la muestra son seleccionados en virtud a

sus posiciones sociales, situación en una red sociométrica, en una cadena de comunicación, en una jerarquía de dominación, etc. Es decir, las muestras estructurales tienen como unidad muestral aquellos elementos que están conectados por una relación específica. Lo más habitual es que no existan listados de individuos en función a sus situaciones relacionales, por lo que parece más apropiado un muestreo intencionado, que aplicar uno probabilístico en este tipo de casos. Un buen ejemplo de este tipo de muestreo es el estudio sobre pautas de formación de grupos informales efectuado por James (1953).

2.9. MUESTREOS CON INFORMANTES ESTRATÉGICOS

Este tipo de muestreo es un compuesto del muestreo homogéneo y del estructural, y parte de la premisa de que el conocimiento y la información está desigualmente distribuido. Resulta de mucho interés el localizar aquellas personas que poseen la mayor información sobre un fenómeno o sistema social.

En este tipo de muestreo basado en informantes, tal como observa Smith (1975), la mejor comprobación de la calidad de los datos obtenidos es la consistencia y repetición de la información que faciliten los datos surgir, y los datos forman un todo coherente, es posible suspenderse la recolección de los datos. Los dos subtipos de muestreo con informantes estratégicos más importantes son el muestreo de “bola de nieve” (*snow ball*) y el muestreo “experto”.

a) Muestreos de “bola de nieve”

Los muestreos de “bola de nieve”, también llamados en cadena, se emplean generalmente en poblaciones especiales o de difícil acceso. Para efectuar el muestreo, se pide a una serie de informantes iniciales que suministren los nombres de otros miembros potenciales de la muestra y que formen parte de la población especial. Este tipo de muestras es en definitiva, un sistema de aproximación para localizar informantes, a partir de las tramas de interrelaciones internas de todo grupo.

b) Muestreos “expertos”

Este tipo de muestreo consiste en pedir a un experto que elija individuos “típicos”, “ciudades representativas”, etc. La investigación sociológica, en especial en norteamérica, abunda de ejemplos de este tipo de muestreos basado en el juicio de expertos. El problema, sin embargo, es que frecuentemente “*los expertos mantienen opiniones diferentes sobre el mejor método para seleccionar casos representativos, o decidir cuál de ellos lo es más*” (Kish, 1965).

Debe recordarse que la lógica de detección de casos típicos apunta a la cualificación e ilustración de perfiles característicos, pero no pretende ni justifica el efectuar generalizaciones a partir de la experiencia de los individuos que componen la muestra.

2.10. MUESTREO DE CRITERIO

La lógica de muestreo es seleccionar para su estudio aquellos casos que se ajustan a algún criterio predeterminado. Por ejemplo, en un estudio sobre conducta desviada, seleccionar aquellos casos con condenas superiores a un número determinado de años, cuando ha sido debido al peso de los antecedentes, aplicando un criterio no basado tanto en la importancia del último delito como de la trayectoria del delincuente. Este tipo de muestreo es especialmente útil en los estudios de evaluación de programas en la medida que, tal como afirma Quinn (1988) *“En el muestreo de criterio la cuestión es estar seguro de comprender bien aquellos casos que probablemente posean mucha información, por que pueden revelar las principales debilidades del sistema”*. Por ejemplo, si el período normal de participación en un programa de salud mental es de 4 a 26 semanas, todos los casos que excedan de 28 semanas deben de ser estudiados para comprobar que son tratados correctamente.

2.11. MUESTREO DE CASOS CRÍTICOS

Un caso crítico es aquel que puede servir de referencia lógica para el resto de la población o parte de ella, en relación al tema de estudio. Esto es importante porque la identificación de un caso crítico depende de reconocer correctamente la dimensión clave que le otorga esa cualidad. En la práctica se emplean diferentes estrategias para la identificación de casos críticos. Una de ellas parte del planteamiento que argumenta *“si esto ocurre allí, sucederá en todas partes”*, o lo contrario *“si no sucede allí, no pasará en ningún otro lugar”*. Otro enfoque alternativo parte de la observación *“si este grupo está teniendo problemas, entonces podemos estar seguros de que la mayoría de los grupos los padecen también”*. Por ejemplo, si los magnates de la bolsa se arrojan por las ventanas en Wall Street, el resto de la sociedad va a tener serios problemas. Los muestreos basados en casos críticos hacen posibles generalizaciones lógicas derivadas del peso de la evidencia que puede llegar a producir incluso un solo caso.

2.12. MUESTREO DE CASOS CONFIRMATORIOS Y CONTRADICTORIOS

Este tipo de estrategia de muestreo intencional es especialmente interesante una vez que el estudio o investigación está avanzado. Cuando ya se ha explorado un fenómeno social y se han observado pautas regulares de comportamiento, puede ser metodológicamente muy importante muestrear específicamente en búsqueda de casos que sirvan para confirmar o para contradecir las conclusiones preliminares.

La lógica que subyace detrás de este tipo de muestras es la búsqueda de nuevas evidencias que corroboren aquello que se a concluido a partir de los datos recolectados. Puede complementar perfectamente un muestreo probabilístico previo. Es importante el buscar tanto casos confirmatorios como contradictorios.

Serán estos casos contradictorios los que buscarán las teorías rivales para criticar las conclusiones a que haya podido llegarse. Por lo tanto, la búsqueda de casos contradictorios e integrarlos en la explicación que se ofrece es fundamental. En ese sentido, se trataría de localizar “*excepciones que confirmen la regla*”, es decir, excepciones que puedan ser correcta y completamente explicados desde las conclusiones que se propone.

Ninguna de las estrategias de muestreo que aquí se han desarrollado son excluyentes entre sí, sino que por el contrario, un buen diseño puede integrar, en la medida que suponga una solución óptima, varios de estos muestreos. En todo caso, debe de mantenerse muy presente que la credibilidad última que se obtenga, para las conclusiones basadas en muestreos intencionados, no depende del tamaño de la muestra ni de la forma de extracción y selección de los individuos, sino de la garantía que ofrezca la riqueza de información que posean los casos considerados.

3. EL MUESTREO PROBABILÍSTICO

El muestreo probabilístico posee una potencialidad muy especial dentro de la investigación científica. Para ello la herramienta fundamental es la estadística inferencial. Tal y como expresa “El propósito principal de los métodos estadísticos es legitimar generalizaciones sobre poblaciones usando datos de muestras”. No es habitual el emplear censos (es decir estudios de la población completa) en los cuales la estadística adecuada es la descriptiva. Ciertamente, una media obtenida de una muestra tiene un margen de error, un intervalo de confianza dentro del cual se encuentra probablemente el valor real de la población. Si los datos proceden de toda la población, la media calculada es “la media”. Posiblemente pueda contener otros errores o deficiencias (no respuesta, falta de comprensión en los contenidos, respuestas sistemáticas, etc.) pero no tiene sentido aplicar la estadística inferencial.

Cuando se aplica la técnica de encuestas a una muestra de la población implica: identificar una población, determinar el diseño de una muestra, seleccionar una muestra representativa, obtener datos de esa muestra, grabarlos y tratarlos estadísticamente para su análisis de forma que puedan efectuarse inferencias sobre el conjunto de la población. Como ya se indicó, la ventaja del muestreo probabilístico reside en su capacidad para estimar el posible error que afecte a la medición (estimado muestral) en su intención de determinar el valor del parámetro poblacional. A cambio de esa potencialidad, el diseño muestral probabilístico está sujeto a fuertes restricciones técnicas, que son las que permite que estadísticamente exista un aval respaldando los resultados de una encuesta. En ese sentido, el muestreo aleatorio simple, por ejemplo, que es el menos complejo teóricamente hablando es, en un sentido práctico, el más difícil de utilizar.

Esto quiere decir que la teoría del muestreo se complicará en tanto en cuanto intente formular estrategias que sean útiles en la práctica para la investigación. Entre una bolsa opaca llena de canicas de colores y las formas de asentamiento territorial y de convivencia de los individuos existen ciertamente diferencias notables.

El hecho básico es que la teoría del muestreo, en sus nociones centrales, es bastante asequible, y las complejidades que llega a adquirir son precisamente consecuencia de la diferencia entre la sociedad real y una situación experimental (por ejemplo, una bolsa llena de bolas blancas y negras). Una estrategia complicada de muestreo probabilístico acostumbra a ser la respuesta a que la realidad social que lo es aún más. Si existiera un ordenador accesible que contuviera el censo de población, y de un modo rápido y económico pudiese extraer muestras aleatorias, los problemas serían exclusivamente de tipo económico para recoger dichas muestras. Cada elemento muestral tendría nombre y dirección, y la extracción sería directa.

Esa situación ideal no existe, y lo que se hace es emplear datos agregados del censo, o del padrón municipal, para el diseño de la muestra y recurrir a mecanismos probabilísticos para la solución final. Queremos, en este sentido concluir con una idea que no por conocida parece ser fácil de aceptar por algunos investigadores: la complejidad de los diseños muestrales responde a problemas reales de investigación. En otras palabras, toda esa complejidad está en definitiva intentando superar las dificultades de selección muestral en determinadas circunstancias.

Lo habitual es que el muestreo esté integrado en una estrategia más amplia de investigación y, evidentemente, hay otras decisiones que influirán en el tipo de muestra que se efectúe. Entre esas decisiones se encuentra el factor económico. El diseño muestral que se elija tiene, generalmente, una gran dependencia de los recursos económicos de que se dispone. Así pues, el diseño muestral forma parte de un proceso de investigación en el que se integra, donde existe un tema de estudio, unas técnicas de aplicación para los instrumentos de recolección de datos y una población sobre las que aplicarlo.

Vamos a considerar el muestreo en dos apartados. El primero está referido a los aspectos donde el investigador debe tomar decisiones sobre las limitaciones que la realidad impone sobre un diseño basado en muestras. Aquí y sobre estos aspectos, el investigador adquiere un claro conocimiento de la capacidad de inferencia de las mediciones que realice. Una encuesta efectuada por Internet solo llega a los que emplean Internet y quieren contestarla. La inferencia es limitada a esa población. Una encuesta telefónica está limitada a los que tengan teléfono y consten en la guía de teléfonos. La relación entre los sujetos alcanzables y la población es una dimensión “cualitativa” y de reflexión por parte del investigador. En un segundo momento trataremos las cuestiones estadísticas, relacionadas

con el diseño formal de la muestra. Aquí las decisiones están muy codificadas, en el sentido de que las preguntas y opciones de respuesta están consideradas de forma sistematizada por la estadística inferencial.

El punto de partida para afrontar un diseño muestral es definir cuál es la población a estudiar. Este concepto de *“población”* es empleado en un sentido técnico indicando la totalidad de los elementos en estudio o universo. Por elementos se entiende la unidad de análisis. La unidad de análisis no tiene por qué estar compuesta por individuos, sino que puede tratarse de otros tipos, por ejemplo instituciones. Si quisiéramos saber la recaudación media de los teatros en España, la población estaría constituida por todos los teatros en funcionamiento en España, donde la unidad de análisis o elemento sería cada teatro. No debe, por lo tanto, confundirse unidad de análisis con individuos, en la medida que una unidad de análisis puede también ser una fábrica, una empresa, un grupo de presión, una ciudad, una región o un estado.

Es fundamental una buena definición de la población de acuerdo a los objetivos del estudio, dado que los resultados que se obtengan dependerán directamente de ello.

Cuando se habla de población general, se entiende que la población sobre la que se efectúa el estudio está compuesta por hombres y mujeres de más de 18 años de edad, que habitan dentro de un área geográfica definida. Este tipo de población general es bastante empleado en los estudios de opinión pública. Otro ejemplo son los estudios sobre la juventud, que acostumbran a emplear un tramo de edad más corto, entre 15 y 29 por ejemplo. Los estudios de fecundidad definen otro tipo de población específica (mujeres en edad fértil), y así un largo etcétera en lo que a definición de poblaciones se refiere. El decidir a qué población se dirige el estudio sirve para tener presente qué elementos forman parte de la población y cuáles no. Esto es más inmediato cuando las poblaciones sobre las que se efectuará el muestreo son específicas, como la juventud, donde el problema consiste en ajustar un tramo de edad socialmente significativo al concepto. Pero en el caso de *“población general”*, hay que recordar que se excluyen los ingresados en hospitales, cárceles, hoteles y cuarteles, así como los que se encuentran de viaje en el extranjero, y determinados segmentos sociales, tanto por clase alta, como baja. En el caso de encuestas a estudiantes, realizadas en el aula, se pueden producir fuertes sesgos, consecuencia de la falta de asistencia de parte de los alumnos. En ese sentido, la población que alcanza y de la que pueda ser representativa la muestra son en la práctica los estudiantes que asisten a clase y no los estudiantes matriculados. Es casi imposible entrevistar en los barrios residenciales de clase social muy alta, tanto como en las barriadas marginales de chabolas. Podríamos seguir enunciando elementos excluidos en la práctica de la población, pero lo importante es evaluar en qué grado su presencia o no presencia puede afectar al estudio. En todo caso, tener siempre presente cuál es la

población real a la que nos referimos. No es exactamente el marco muestral (de forma simple, listado del que se extrae los elementos de la muestra, como por ejemplo la guía telefónica), en la medida que en este listado pueden existir individuos que son en la realidad inaccesibles.

Cuando la población ha sido definida, es el momento de afrontar la tarea de obtener una muestra de ella. La opción de tomar todos los elementos de la muestra se denomina censo, y su realización es excesivamente costosa para que se pueda hacer con la frecuencia que requieren los estudios de opinión. Precisamente, la alternativa a tomar todos los elementos de la población y seleccionar sólo una parte de ellos es lo que se denomina muestreo.

El muestreo no sólo es más factible económicamente que realizar un censo, sino que posee además las ventajas de que puede prestarse una mayor atención a la recolección de datos, elevando así la calidad de éstos. Al mismo tiempo, dicha recolección, así como el proceso de la información y la elaboración del informe es mucho más rápido. Todas estas ventajas compensan sobradamente la sustitución del censo por una muestra, siempre y cuando los estimados posean una fiabilidad razonable. Precisamente, el objetivo de la teoría del muestreo es obtener estimados lo más próximos a los valores existentes en la población.

La cuestión central de cualquier diseño muestral es cómo seleccionar aquellos elementos de la población que van a formar parte de la muestra. Es decir, la extracción del elemento (individuo, institución, etc.) que va a facilitar la información. Si el criterio de selección es probabilístico, cada elemento de la población posee una probabilidad conocida y diferente de cero para ser incluido en la muestra. La consecuencia de esto es que se evitan los sesgos de selección y es factible emplear la teoría estadística para considerar las propiedades de los estimados.

Una vez definida la población que va a ser objeto de muestreo, necesitamos alguna forma de identificar a los elementos que forman parte de la población. A esto se le denomina “*marco muestral*”. Deming (1960) define el “*marco muestral*” como “*el conjunto de materiales físicos (censos, mapas, listas, directorios, archivos) que nos permite dar cuenta del universo pieza a pieza*”. Sin embargo, el “*marco muestral*” está en relación al diseño de muestra que se elija. Así, en un muestreo polietápico, es evidente que la lista que se posea en la primera etapa no será la de las unidades de análisis que al final son el objeto del estudio. En un muestreo de área “*toda el área en el que se localiza la población se subdivide en áreas más pequeñas, y cada unidad elemental es asociada con una, y sólo una, de esas áreas*” Hansen (1953). Existen dos formas de construir un “*marco muestral*”, que coinciden con los dos modos de definir en lógica un conjunto (o subconjunto). El primero de ellos consiste en hacer u obtener un listado con todos los elementos. El segundo, proveer de una regla para identificarlos. Listar los elementos no es siempre posible y muchas veces es una opción más positiva, en la medida en que los elementos puedan ser identificados, el generar una regla

para localizar y seleccionar los casos. En el caso que se posea un listado con todos los elementos de la población, ese constituirá el “*marco muestral*”. El poseer un buen “*marco muestral*” que facilite como mínimo de un medio para identificar y localizar los elementos, así como el modo en que este organizada esta información, tiene una influencia sustancial en el diseño muestral. Un “*marco muestral*” organizado por áreas geográficas lleva directa mente a un muestreo por “*agrupamiento*”, mientras que si están organizadas por factores relevantes para la investigación puede sugerir un muestreo estratificado. En todo caso, la calidad de la información que constituye el “*marco muestral*” es esencial para un muestreo eficaz y fiable. Los problemas más frecuentes que pueden aparecer en un “*marco muestral*” son los cuatro considerados por Kish (1965): elementos perdidos, elementos extraños, agrupaciones y duplicaciones.

La situación por la que determinados elementos no aparecen en el “*marco muestral*” puede deberse a que éste sea inadecuado. Es decir, que la intención con la que se genera ese “*marco muestral*” no pretendía abarcar a toda la población que va a ser objeto del estudio o porque este sea incompleto, ya que pretendiendo cubrir toda la población falla en ese intento. Las dos pueden ser causa de elementos perdidos, debiendo analizarse detalladamente tanto si el “*marco muestral*” es el adecuado, como su completitud. De todos modos, es más fácil identificar un “*marco muestral*” inadecuado que uno incompleto, y de hecho las posibles soluciones son más accesibles para el primer caso. Los elementos perdidos constituyen un problema bastante serio pues esos casos no poseen ninguna probabilidad de estar incluidos en la muestra, con lo que ésta dejará de ser representativa de la población.

Algunas veces es posible evitar el problema de los elementos perdidos redefiniendo la población de forma que los excluya. Esto es especialmente fácil de hacer en el caso de los “*marcos muestrales*” inadecuados, aunque es evidente que constituye una solución bastante imperfecta. Se debe intentar completar el “*marco muestral*” reuniendo a otras posibles fuentes alternativas, si bien esto nos llevara al problema de la posible duplicación de elementos.

Los elementos pueden aparecer duplicados, por encontrarse en diferentes listados, lo que incrementa su probabilidad de ser seleccionado para la muestra. Este es un riesgo serio al emplear diferentes fuentes secundarias. Otro problema importante es cuando los posibles listados que están disponibles presentan datos excesivamente agregados, con lo que obstaculizan el problema de afijación y selección de la muestra. Por último, los listados o datos estadísticos pueden recoger elementos que no estén incluidos en la población objeto del estudio, con lo que afectarán al proceso de diseño muestral.

Se debe de intentar la mayor coincidencia posible entre el “*marco muestral*” y la población objeto de estudio, dado que, estrictamente hablando, sólo puede efectuarse inferencias sobre elementos presentes en el “*marco muestral*”.

Un aspecto importante es el de la homogeneidad de la población, siempre con respecto a la característica que es objeto de estudio. Por ejemplo, si queremos saber si el mar es salado o no, ¿cuántas veces habrá que mojar el dedo en el? Evidentemente, para responder a la pregunta basta con mojar una sola vez el dedo. Para una población tan grande como el mar, basta con una muestra compuesta por solo un elemento, si lo que deseamos saber es su sabor. Si deseamos saber el nivel de salinidad según profundidad, necesitaremos muestrear a diferentes profundidades. La pregunta es la que guía el diseño y será la que exija una mayor control de heterogeneidad.

En general, las diferentes estrategias de muestreo intentan optimizar el conocimiento de la población en estudio para conseguir una información más fiable con el menor coste. Como ya advertimos, el necesario e imprescindible respeto a la aleatoriedad en el procedimiento de extracción muestral, convierte en muy costosa y difícil toda la tarea del personal de campo. Este es, no obstante el precio a pagar por conocer aproximadamente el sesgo que se ha introducido en el estimado muestral.

En definitiva, el principio a respetar es que la muestra sea representativa, lo que significa que todos los individuos que componen la población tienen una probabilidad conocida e independiente de formar parte de la muestra. En una muestra aleatoria, y por lo tanto representativa, la probabilidad de seleccionar a un individuo es la misma que la de seleccionar a cualquier otro. En segundo lugar, la probabilidad de seleccionar a un individuo debe de ser independiente de que se elija a cualquier otro. Es decir, la extracción de un individuo no condiciona las probabilidades de que otro sea elegido. En la práctica, un muestreo aleatorio puede operativizarse de varias formas. La más tradicional es la utilización de tablas de números aleatorios. Aun cuando parezca paradójico, una tabla de números aleatorios es una lista de dígitos “ordenados” aleatoriamente del 0 al 9.

Vamos a considerar un ejemplo. Supongamos que tenemos la lista de alumnos (definidos según cursar una materia, o que pertenecen a un grupo, etc.). Primero numeramos a los alumnos. Es decir, asignamos un número a cada uno de ellos. Después tomamos una tabla de números aleatorios y decidimos en qué dirección la vamos a leer (de izquierda a derecha, de derecha a izquierda, comenzando arriba hacia abajo, en diagonal, etc.). Después se toma un número de inicio cualquiera y se comienza a tomar números aleatorios sucesivos, tal y como aparecen. Deben tomarse tantos dígitos en cada número como sea significativo para el tamaño de la población. Por ejemplo, si los alumnos de la clase están numerados de 1 a 70, y el número aleatorio es, por ejemplo, 137654, tomaríamos solamente las dos primeras cifras (13) y pasaríamos a las cifras siguientes (76), (54), etc. No importa seleccionar más números de los que se necesitan para la muestra o números que exceden la población o números repetidos. Ciertamente, para seleccionar la muestra se emplearán los números significativos

(menores de 71 en este ejemplo) y no repetidos. Los espacios en blanco se introducen para facilitar la legibilidad de los números, de forma que su agrupación en columnas es arbitraria y debe ignorarse en la secuencia de extracción del siguiente elemento muestral.

Una alternativa a la extracción aleatoria anterior, es la extracción mediante procedimiento sistemático. No obstante, no existe una teoría estadística que legitime de manera precisa la exactitud de las estimaciones mediante este procedimiento. En la práctica, se las trata como si fuesen muestras aleatorias. Como podemos apreciar, la aleatoriedad, es decir, la posibilidad de participar en la muestra, debe mantenerse desde el diseño inicial, hasta la extracción misma de la muestra. En el caso de las encuestas de opinión pública, la aleatoriedad en la extracción final se intenta garantizar mediante los sistemas de rutas aleatorias, donde el entrevistado (ante la imposibilidad evidente de mezclar a la población) se mezcla según criterios de aleatoriedad (eligiendo calles, eligiendo portales, eligiendo viviendas, eligiendo al entrevistado dentro de la vivienda normalmente según tabla de números aleatorios). Existen varios procedimientos que aplican en la práctica la opción de rutas aleatorias. En general, la idea es que los agrupamientos de individuos (colas en el mercado, en el cine, etc.) pueden contener algún sesgo en el sentido de acentuar la homogeneidad. Por ello, la fragmentación de la extracción donde los individuos se seleccionan uno a uno, aplicando algún criterio de “mezcla” entre cada entrevistado.

Considerando la estadística que existe detrás de la teoría del muestreo encontramos las ideas clásicas de la probabilidad. El muestreo se apoya fuertemente en la idea de las repeticiones a largo plazo. Por ejemplo, que si se arroja una moneda al aire muchas veces, saldrá cara un 50% de las veces. Y de forma consecuente: si extraes un número elevado de muestras de una población, a largo plazo las muestras mostrarán representatividad de la población. Así, la cuestión se reduce a obtener una muestra aleatoria y determinar la probabilidad de que reproduzca las mismas características que la población de la que ha sido tomada. El valor que pueda poseer un coeficiente en la población (una media, una proporción, una desviación, etc.) se denomina “parámetro”, mientras que el obtenido mediante una muestra se le llama “estimado”.

La confianza en que un estimado esté próximo al parámetro se incrementa con el tamaño muestral. En el extremo, una muestra que contenga el 100% de la población (sería un censo) tendría una confianza del 100% respecto a contener las características de la población. En general, decidiremos qué confianza queremos antes de diseñar la muestra. Porcentajes de confianza del 99%, del 95% o del 68%, aproximadamente, son los que habitualmente se emplean como referencia (al coincidir con las desviaciones típicas en la distribución normal). Pero, obviamente, el investigador decide cuál es la confianza requerida. Una confianza del 60% puede ser suficiente si existen razones para ello.

La población puede considerarse como infinita (en la práctica más de 100000 casos) o finita (menos de 100000). Según sea el tamaño de la población, se pueden emplear correcciones en el cálculo de la muestra. Tal y como afirmábamos, la confianza crece cuanto mayor es el tamaño de la muestra. Con esto nos estamos refiriendo a la muestra exclusivamente y no a su relación con la población a la que se refiere y de la que ha sido extraída, también conocida como fracción de muestreo. La razón del tamaño de la muestra, n , al tamaño de la población, N , se conoce como la fracción de muestreo, n/N . Para poblaciones finitas, a menos que la fracción de muestreo exceda .05, pueden emplearse sin diferencias significativas los mismos criterios que cuando N es infinita. En ese sentido, en lo que se refiere al tamaño de la muestra, es éste el que importa y no tanto la fracción de muestreo.

Otro procedimiento factible para elevar el nivel de confianza es relajar la precisión del estimado. Así, consideremos el siguiente ejemplo. Podemos afirmar que la media de autoubicación ideológica (1 a la izquierda y 10 a la derecha) es de 4,5, con un intervalo de confianza (es decir, que el parámetro o valor real de la población se encontrara contenido dentro del intervalo) entre 4,2 y 4,8. También podemos relajar la precisión y afirmar que la media es de 4,5 con un intervalo de confianza de 3,5 a 5,5. La amplitud del intervalo, es decir la precisión de la medición efectuada, está en relación con el nivel de confianza. Es decir, podemos afirmar que con una confianza del 60% el parámetro (valor en la población) esta dentro de un intervalo entre 4,2 y 4,8 o por el contrario subir el nivel de confianza y abrir más aún el intervalo: con una confianza del 75% el valor real está dentro de un intervalo entre 3,5 y 5,5. Nuevamente, llevando la lógica al extremo, con un nivel de confianza del 100% el valor real se encontrará entre 1 (izquierda) y 10 (derecha).

Es decir, a efectos de diseño de la muestra es importante recordar que existe una relación fija entre entre tres conceptos: el tamaño muestral, el nivel de confianza y el intervalo de confianza. Si se determinan dos de ellos, el tercero está fijado por defecto.

El fundamento estadístico de la teoría de muestreo se sustenta en el Teorema del límite central, que afirma: independientemente de la forma de la distribución en la población de origen, la distribución que forma las medias estimadas a partir de un número elevado de muestras independientes, se aproxima a la normalidad conforme n (el tamaño de cada muestra) se incrementa. Es decir, si extraemos 1000 muestras diferentes de una población, las medias que se estimen formarán una distribución normal. Más rápidamente adoptará esta forma cuanto mayor sea el tamaño de las muestras que se extraen. Este fenómeno se conoce en estadística como el *teorema del límite central* donde la distribución de medias de la muestra (X) se aproxima a una distribución normal conforme el tamaño de la muestra, n , se incrementa. El teorema del límite central ha sido llamado “el

teorema más importante en estadística desde el punto de vista teórico y aplicado” (Snedecor y Cochran, 1980) y “uno de los teoremas más notables en todas las matemáticas” (Mood y Graybill, 1963).

Es muy importante apreciar que este teorema fundador de la teoría del muestreo no afirma nada sobre la distribución o las observaciones en la población de la que se extrae la muestra. Independientemente de la distribución en la población, la forma de la distribución de muestreo de la media se aproxima a la distribución normal si n es lo bastante grande. Una vez explicitada la regularidad, debemos tener en cuenta las características de la distribución en la población para determinar el tamaño muestral suficientemente grande para dar sustento al teorema. El tamaño adecuado de las muestras para obtener la normalidad en las distribuciones de estimados, depende de la forma de la distribución de la población origen. Solamente en el caso de distribuciones muy atípicas, un tamaño muestral de 25 generará una distribución de estimados próximos a la distribución normal. Incluso para poblaciones de origen no normales, la forma de las distribuciones de muestreo rápidamente se aproximan a la normalidad a medida que el tamaño muestral aumenta. Además, conforme el tamaño muestral se incrementa, la variabilidad de las distribuciones de muestreo disminuye incluso si la población de origen no es normal.

Gracias a esta propiedad es posible cuantificar el error que pueda contener un estimado en relación al valor real en la población. Cuando el tamaño muestral se incrementa, la distribución que adoptan los estimados muestrales (la media de cualquier cosa) se aproxima a una distribución normal independientemente de la forma que tenga en la población de origen. Es importante apreciar la idea de repetir la extracción de múltiples muestras, cada una de ellas con su propia media.

Es bien conocido que si en una distribución normal sabemos su media y la desviación típica, sabemos todo lo que necesitamos. En ese sentido, al efectuar un muestreo vamos a considerar la relación entre la media que nos ofrece nuestra muestra con la media de la distribución de muestras que sabemos coincide con el valor real en la población. A partir de esta relación, se derivan toda una serie de consecuencias muy útiles y especialmente la que conduce al nivel de confianza, la construcción de intervalos de confianza y la cuantificación de un error estimado.

Aproximadamente el 95% de los valores en una distribución normal se encuentran entre dos desviaciones estándar de la media. Una muestra grande de medias (procedentes de varias muestras) estará normalmente distribuida y su desviación estándar se denomina error estándar de la media. Así, existe un 95% de probabilidad de que la media estimada en una muestra cualquiera se encuentre dentro de un error estándar de la media de 1,96 (es decir, valor z expresando el número de desviaciones estándar respecto a la media) alrededor de la media

(parámetro o valor real) de la población. Es decir, tendremos un 95% de confianza en que, media estimada de una muestra cualquiera = media de la población $\pm [1,96 \times (\text{desviación estándar en la población} / \sqrt{\text{del tamaño muestral}}]$ o planteando al revés esta relación media de la población = media estimada de una muestra cualquiera $\pm [1,96 \times (\text{desviación estándar de la muestra} / \sqrt{\text{del tamaño muestral}}]$

En la práctica esto es muy simple de aplicar. Supongamos que tenemos una titulación y extraemos una muestra con las calificaciones de 50 alumnos y estimamos que la calificación media es de 6,7 con una desviación estándar de 2,3. Podríamos decir con un 95% de confianza que,

Nota media de todos los alumnos de la titulación = $6,7 \pm [1,96 \times (2,3/\sqrt{50})]$

luego podemos afirmar con un 95% de confianza que la nota media de la titulación se encuentra dentro de un intervalo entre $6,7 \pm 0,63$; es decir, entre 6 y 7,3

Como ya advertíamos, los intervalos de confianza dependen del nivel de confianza. El nivel de confianza viene expresado en puntuaciones Z (dispersión entorno a la media). Algunas puntuaciones Z importantes y su nivel de confianza son:

Puntuación Z	Nivel de confianza (Área entre las colas)
1	68,26
1,64	90
1,96	95
2	95,4
2,58	99

En general, aplicaremos la fórmula ya presentada, si bien variaremos el valor Z en función a la confianza que se desee.

Media parámetro (población) = media estimada (muestra) $\pm [Z \times (\text{desviación típica estimada} / \sqrt{\text{tamaño muestral}}]$

Esta relación nos permite responder a la mayoría de las cuestiones relevantes. En primer lugar recordemos que el error que obtenemos en nuestra muestra es simplemente la diferencia entre la media en la población (parámetro) y la media en la muestra (estimado). Recordemos que el error es una de las referencias para construir el intervalo de confianza. Es decir, el error es igual a:

Error = $Z \times (\text{desviación estándar} / \sqrt{\text{tamaño muestral}})$

En el ejemplo anterior,

Nota media de todos los alumnos de la titulación = $6,7 \pm [1,96 \times (2,3/\sqrt{50})]$

Parámetro – estimado (6,7) = Error = $1,96 \times (2,3 / \sqrt{50}) = 0,63$

Y para el cálculo del tamaño muestral despejamos nuevamente

tamaño muestral = $(Z \times \text{desviación estándar} / \text{error})^2$

en nuestro ejemplo

tamaño muestral = $(1,96 \times 2,3 / 0,63)^2 = 51$

Recordemos que la muestra original era de 50 casos. La diferencia es consecuencia del redondeo durante las operaciones. En general, para determinar un tamaño muestral, deberemos considerar el nivel de confianza (Z), una estimación de varianza en la población y el error que nos deseamos permitir. El tamaño muestral que decidamos tener depende de cuatro factores. En primer lugar del número de subgrupos que deseamos analizar. Así, si se desea analizar según género, la muestra global quedará reducida a prácticamente la mitad para muchos análisis. El segundo factor es el valor que tenga la información que queremos recoger y por lo tanto el grado de error que nos podemos permitir. El tercer factor es el coste de la muestra. Si el coste es bajo, quizás sea permisible un tamaño muestral elevado. El cuarto y último factor es la variabilidad que previsiblemente contenga la población. Si todos los individuos mantienen la misma opinión, una muestra de uno será suficiente. Si por el contrario, todos tienen una opinión diferente (por ejemplo el número de DNI), ninguna muestra será lo bastante buena para capturar esa variabilidad extrema.

El concepto y la lógica que se utiliza en el caso de las medias es aplicable a cualquier otro estadígrafo como proporciones, dispersión, etc. Las proporciones en el caso de distribución binomial son esencialmente medias disfrazadas. Por ejemplo, cuando una variable está codificada con 0 y 1, la media es realmente la proporción de 1 en la distribución. Cuando se trata de estimar proporciones (ejemplo, el porcentaje de individuos que expresa una opinión concreta), aplicaremos el mismo procedimiento, solo que modificaremos la estimación de variabilidad. Sustituiremos la desviación estándar por la proporción esperada para algunas de las opiniones.

Vamos a sustituir la desviación estándar por una estimación de varianza, mediante la relación $p(1-p)$, donde $1-p$ suele denominarse q . Recordemos que la varianza es la desviación estándar elevada al cuadrado.

tamaño muestral = $Z^2 \times p(1-p) / \text{error}^2$

Es habitual emplear la situación más desfavorable donde se produce una indeterminación del 50%. Así, tendríamos la situación siguiente. Supongamos que queremos saber el tamaño muestral para saber qué porcentaje de adultos fuma en España, con una confianza del 95% y un error del 5% en la situación más desfavorable (p/q 50/50). Para el nivel de confianza emplearemos el valor

Z, para el error lo expresaremos en tanto por 1 en lugar de en tantos por ciento (un error del 5% es un error del 0,05 por 1).

$$\text{tamaño muestral} = 1,96^2 \times 0,25 / 0,05^2 = 384 \text{ casos}$$

Y viceversa, si lo que deseamos saber es el error, procedemos como se hizo anteriormente. Nuevamente recordemos que la varianza es la desviación estándar elevada al cuadrado. Por ello, obtenemos la raíz cuadrada de la varianza.

$$\text{Error} = Z \times (\text{desviación estándar} / \sqrt{\text{tamaño muestral}})$$

$$\text{Error} = Z \times (\sqrt{\text{varianza}} / \sqrt{\text{tamaño muestral}})$$

$$\text{Error} = 1,96 \times (\sqrt{0,25} / \sqrt{384}) = 1,96 (0,5 / 19,59) = 0,05$$

Como puede apreciar, los procedimientos para determinar el intervalo de confianza, el nivel de confianza, el error o el tamaño muestral son bastante evidentes, en el caso del muestreo aleatorio simple aquí considerado.

Por último, es importante efectuar algunas consideraciones relevantes sobre la práctica del muestreo. En general, podemos confiar en que la presunción de la distribución normal va a funcionar con muestras de aproximadamente 30 o más casos. Para muestras inferiores a 30 casos podremos, con cierta confianza, emplear los valores Z cuando consideremos que la distribución en la población es aproximadamente normal siempre que conozcamos la desviación típica de la población. En el caso que empleemos la desviación estándar procedente de la muestra en lugar del parámetro, obtendríamos una distribución platicúrtica. En el caso de muestras menores de 30 casos es más interesante emplear la distribución de la de t de Student. Simplemente utilizar los valores de t en lugar de los valores Z. Otro problema es cuando el tamaño muestral es mayor que el 5% de la población. Cada vez que se extrae un elemento muestral el tamaño de la población se reduce. Por lo tanto, la probabilidad condicional de elegir un nuevo elemento muestral es mayor que para el elemento extraído anteriormente. Para corregir esto multiplicaremos el error estándar de la media por el factor de corrección de poblaciones finitas, donde n es el tamaño muestral y N el tamaño de la población ($\sqrt{[(N - n) / (N - 1)]}$). Muchos programas ignoran este factor de corrección en la presunción de que la población originaria es siempre grande o infinita. Por otra parte, es interesante incluirlo en todos los cálculos, en la medida en que su efecto tiende a suprimirse cuando el tamaño de la población se incrementa en relación al tamaño de la muestra. Estas consideraciones que son válidas para el caso de las estimaciones de medias, son trasladables al caso de la estimación de proporciones. La distribución binominal sólo se aproxima a la normal cuando n (tamaño muestral) es mayor que 30 y tanto $n \times p$ y $n \times (1 - p)$ son mayores que 5. Ciertamente, los estadísticos proponen que cuando la pro-

porción (p) no es conocida, la distribución normal no deberían emplearse como referencia con muestras menores de 100. De ser así, el intervalo de confianza debería ser calculado con una proporción fija al 0,5, lo que produce los estimados más conservadores posibles. Al igual que en el caso de la estimación de medias, cuando el tamaño muestral es mayor que el 5% de la población, el error estándar de la proporción deberá ser multiplicado por el factor de corrección finita que se mencionó anteriormente.

BIBLIOGRAFÍA

- ALAMINOS, A. (2003) *El análisis de la realidad social*. Alicante Obets.
- ALAMINOS, A. (1994) *La encuesta en los países en vías de desarrollo. Teoría y práctica*. CEDEAL.
- BARKER, STEPHEN F. (1974) *Elements of Logic*. NY: McGraw-Hill.
- ARTHUR BOWLEY y A.R. BURNETT-HURST (1915) *Livelihood and Poverty*.
- CUBER, J. & HAROFF, P. (1968) *Sex and the Significant Americans: A study of sexual behaviour among the affluent*. Penguin.
- DEMING, WE. (1960) *Sample Design in Business Research*. Wiley and Sons.
- DENZIN, N. K. (1970) *Sociological Methods: a Source Book*. Aldine Publishing Company. Chicago.
- HANSEN, MH; HURWITZ, WN; MADOW, WG; (1953) *Sample Survey Methods and Theory* John Wiley and Sons, New York, Vols. Iy II.
- HOPKINS K., HOPKINS B., y GLASS G. (1997) *Estadística para las Ciencias Sociales y del Comportamiento*. Prentice-Hall,
- KISH L. (1965) *Survey Sampling*, New York: John Wiley & Sons.
- MOOD, A., GRAYBILL, F. and BOES, D. (1963) *Introduction to the theory of Statistics*. Wiley, N. York.
- SMITH, H. W. (1975) *Strategies of Social Research. The methodological imagination*. Prentice Hall. London.
- SNEDECOR G.W. and COCHRAN. W. (1980) *Statistical Methods*. Iowa State University Press, seventh edition,
- QUINN M. (1980) *Qualitative Evaluation Methods*, Beverly Hills, CA: Sage Publications,. Inc.,

TEMA 3

LA DINÁMICA DE LA INTERACCIÓN

Antonio Alaminos

1. INTRODUCCIÓN

Uno de los criterios básicos es que las entrevistas deben de efectuarse, idealmente, estando presentes solamente el entrevistador y el entrevistado, o el entrevistado y el cuestionario. No es siempre posible, no solamente cuando una entrevista se efectúa en público, sino también cuando están presentes los miembros de la familia. La razón es que uno de los objetivos primordiales de la introducción del cuestionario en la recolección de datos, especialmente en la encuesta, es el de normalizar al máximo la interacción entre el entrevistador y el entrevistado. Existen otras muchas ventajas indiscutibles, como es el de facilitar la labor de los entrevistadores o el tratamiento posterior de la información. Sin embargo, la mayor potencialidad del empleo del cuestionario es el “*efecto de subrutina*” que produce en todo el proceso de la recolección de datos. El cuestionario ofrece la posibilidad de que todos los entrevistados sean interrogados exactamente con las mismas palabras, presentándose las preguntas en el mismo orden. Es ésta esencialmente la garantía que hace equiparables las respuestas de los entrevistados.

En ese sentido, es evidente el énfasis que debe ponerse en la formación de los entrevistadores, dado que el cuestionario intenta regular la interacción entre estos y los entrevistados. Sin embargo, en primer lugar, será el grado de entrenamiento de los entrevistadores, para atenerse al cuestionario, la garantía última de estandarización.

Esta preparación de los entrevistadores para administrar un cuestionario es, ciertamente, muy diferente a la de los entrevistadores destinados a efectuar

entrevistas no estructuradas. El entrevistador formado para la aplicación de cuestionarios intenta neutralizarse. Para él, un cuestionario es el cauce de la interacción, intentando restringir y controlar la información que se facilita al entrevistado a lo que ya está escrito. Una diferencia de personalidad en los entrevistadores, sobre todo si el número de ellos es reducido, puede, de ser dejados a su libre albedrío sin orientación o formación previa, introducir efectos muy fuertes en la información final que se recoge.

Este es un aspecto esencial de la encuesta: la normalización de las interacciones que se desarrollan durante la administración del cuestionario es la garantía fundamental que se tiene para poder hacer comparables las respuestas de todos los entrevistados. O lo que es lo mismo, que las diferencias que se encuentren en las respuestas correspondan con diferencias de la población y no con diferencias de los entrevistadores. Veamos esto en más detalle.

Supongamos un caso teórico con solo dos entrevistadores A y B, administrando un cuestionario a una muestra seleccionada de la población. El procedimiento que seguirá la información es canalizado a través de estos entrevistadores. Una vez que los entrevistadores, lo suficientemente adiestrados, han aplicado el cuestionario a una muestra probabilística representativa de la población, se obtiene una información proveniente de dicha muestra. Supongamos un análisis efectuado sobre dicha información. En él descubrimos la existencia de dos segmentos, extrapolables al resto de la población. En este proceso, garantizada la estandarización de la recogida de datos, mediante el empleo del cuestionario y el adiestramiento de los entrevistadores, suponemos que su actuación acaba con la devolución de los cuestionarios, convertidos en datos, y desaparecen en el resto del proceso. Esta es la situación ideal que se persigue con un buen personal de campo. Dado un buen adiestramiento de los entrevistadores, la posibilidad de un efecto de éstos sobre los datos obtenidos puede suponerse, de producirse, que se compensan entre ellos.

Existe, por supuesto toda una gradación de posibilidades en estos efectos de los entrevistadores en la información final que estamos considerando. Veamos, por ejemplo, la otra posibilidad extrema, el que los entrevistadores, careciendo de un buen adiestramiento, interfieran en la información final que se obtenga. De nuevo dos entrevistadores (o grupos de entrevistadores de diferente sexo, raza, etc.) con una personalidad marcada o insuficientemente adiestrados. Seguido el mismo proceso que en el caso óptimo anterior, obtendríamos una segmentación extrapolable a la población. Sin embargo, en este caso sí se ha producido un efecto de los entrevistadores sobre los entrevistados, y los segmentos detectados recogen realmente las diferencias existentes entre los entrevistadores y su efecto, y no diferencias reales de la población. Se produce un ciclo cerrado sobre sí mismo, donde el análisis final nos habla del proceso de recolección de datos que le sustenta y no de la sociedad sobre la que estamos trabajando. Este caso, donde

el investigador trabaja sobre los efectos de los entrevistadores y no sobre información extraída de la sociedad es el peor de los imaginables.

Por ello, es especialmente importante la identificación del entrevistador dentro del cuestionario. Nos ofrece la posibilidad de controlar que las diferencias que encontremos en los datos corresponden con diferencias que existen en la realidad y no guardan dependencia con los entrevistadores. Poder afirmar esto es la finalidad última de los intentos de reducir los posibles efectos de la interacción del entrevistador en el entrevistado, mediante la introducción del cuestionario y con una buena preparación de los entrevistadores. Así, por ejemplo, cuando se insiste que las preguntas deben de leerse tal y como aparecen en el cuestionario, y no variar espontáneamente la redacción de la pregunta, es para evitar la posibilidad de que las diferentes redacciones de las preguntas produzcan efectos diferentes en los entrevistados, haciendo así incomparables las respuestas que se obtengan. Como ya hemos afirmado, se trata de garantizar que las diferencias detectadas en los datos correspondan con diferencias reales de la población, y no con diferencias en la recogida de la información.

Un ejemplo de esto lo ofrece Ralis, Suchman y Goldsen (1958), cuando comprobaron la fiabilidad y la validez de los datos obtenidos en una encuesta. El análisis de consistencia interna ofreció la conclusión de que los datos eran válidos y significativos. Sin embargo, se detectó el error introducido por los entrevistadores al contrastar los datos recogidos por el grupo de entrevistadores que eran empleados del gobierno y aquel otro que no lo era, lo que sugería un fenómeno del segundo tipo. La etapa de la recolección de datos es fundamental y ningún tiempo empleado en ella puede considerarse tiempo perdido, dado que condiciona, y determina, tanto la información que se obtenga como las conclusiones que se extraiga de ellas mediante análisis.

Los intentos de normalizar la interacción, bastante lograda en el caso de la encuesta, es esencial en la medida que garantiza la comparabilidad posterior de los datos. Otras, como el empleo prudente de tarjetas ante la posibilidad de encontrar segmentos analfabetos, es un problema casi de sentido común para cualquier investigador social, y previsible en la etapa de diseño de la investigación. La selección y adiestramiento de entrevistadores es, sin embargo, un problema muy específico de cada país, donde las soluciones son menos generalizables. La existencia del problema es absolutamente evidente, donde en ocasiones normalizar el efecto de los entrevistadores se presenta especialmente difícil por la propia heterogeneidad de la población en estudio.

Es necesario un análisis detallado previo de la estructura y características de la sociedad a estudiar, no solo para calcular las estrategias de diseño muestral, sino también para la mayoría del resto de las etapas del proceso de investigación, como vamos a tener ocasión de observar.

Los entrevistadores deben ser adiestrados, pero antes de ello existe la necesi-

dad de seleccionarlos. Esta selección debe de hacerse de acuerdo a las características del tipo de campo que se vaya a realizar, en que contexto cultural y del tipo de materia que vaya a tratarse. Junto a ello, es conveniente que los entrevistadores reúnan algunas cualidades personales especialmente necesarias para el trabajo de entrevistador.

El problema que supone la necesidad de normalizar la entrevista, comienza a dejar de serlo desde el momento en que se le reconoce. Es evidente para cualquiera, que la calidad final de los datos que se recojan en una encuesta depende, en último extremo, de la habilidad de los entrevistadores para realizar las entrevistas. Por ello, un aspecto vital es el de la selección del personal de campo.

El empleo de entrevistadores pertenecientes a las sociedades en las que se va a efectuar el estudio no aporta ninguna solución definitiva en los casos donde predominan las características rurales. Hursh-Cesary Roy (1976) indican, por ejemplo, con respecto al empleo de estudiantes; *“dado que son gente educada, pueden ser contratados con un coste bajo, frecuentemente expresan su interés en la investigación social, y pueden encontrarse fácilmente una gran variedad de subgrupos culturales, se emplean con mucha frecuencia como entrevistadores a los estudiantes de universidad (.../...) Frecuentemente con resultados desastrosos. Guste o no, la mayoría de los estudiantes forman parte de la élite, son ya o se van a convertir en miembros privilegiados de las clases urbanas. Su empleo puede provocar serios problemas de comunicación. Pueden tener grandes dificultades para abandonar sus roles de status social alto y adoptar el comportamiento y estilo que se requiere para establecer conversación con agricultores analfabetos que se encuentran en un nivel de subsistencia”*.

Este problema es especialmente grave en ciertos estudios, como los de pobreza, donde es de primordial importancia conocer el contexto local y controlar, con sumo detalle, el rol social del entrevistador. No hay que descuidar en el momento de la selección de los entrevistadores que estos han de tener un nivel educativo que respalde el que van a responder al adiestramiento, van a entrevistar de un modo responsable y que serán capaces de comprender los objetivos de la investigación. Éste nivel educativo debe de evaluarse en relación a las características de la población. Un nivel educativo excesivamente elevado puede claramente obstaculizar las entrevistas en zonas rurales, al incrementar las diferencias de status entre el entrevistador y el entrevistado. Parece lógico pensar que en zonas urbanas sea más positivo el emplear entrevistados con un nivel educativo diferente a los que administren los cuestionarios en zonas rurales. En general, deben identificarse ciertas características para reclutar a los entrevistadores, con el objeto de reducir las diferencias culturales y de status, facilitando una comunicación fácil con los entrevistados. En este intento de educación de las características de los entrevistadores a los diferentes tipos de interacción social que van a encontrar, es útil considerar el problema de la lengua. En zonas donde se habla

más de una lengua, o existen grupos étnicos, puede ser útil y deseable el seleccionar entrevistadores que reproduzcan esa diversidad étnica y que posean competencia lingüística para hablar las lenguas que se pueden encontrar. Hursh-cesar observa, desde su experiencia en este tipo de ensamblajes entre el personal de campo y la sociedad que *“La mayoría de los entrevistadores encuentran extremadamente difícil explicar su trabajo como recolectores objetivos de datos a los campesinos, que no poseen ningún esquema de referencia significativa en el que ubicar la entrevista realizada en una encuesta. La tarea es más fácil cuando los entrevistadores y los entrevistados son parecidos en términos de origen social, religión, lengua y dialecto, área geográfica y tienen la capacidad de comprender, apreciar y observar costumbres comunes. Compartir esas características reducen barreras a la comunicación y se puede prever que mejore la aceptación y la confianza”*. Parece lo suficientemente lógico el argumento, y en la medida que pueda ponerse en práctica, el consejo es bastante válido. Hay, sin embargo, algunas matizaciones que hacer. Existe la idea de que puede ser conveniente el emplear como entrevistadores a individuos provenientes de los pueblos o lugares en los que se va a entrevistar. No parece ser eficaz este método. En primer lugar, por la dificultad de realizar un adiestramiento correcto de los entrevistadores de un modo puntual, cuando no han tenido ningún contacto previo con la investigación social. En segundo lugar, por la escasez de tiempo con que se cuenta habitualmente en una encuesta. Y, por último, por el hecho de que a veces se encuentran reticencias a ser entrevistado por personas que son conocidas y que poseen una posición social definida dentro de la comunidad, de modo que intervendrían las relaciones de amistad-enemistad al intervenir el entrevistador en un entorno “doméstico”. Quizás la mejor alternativa puede ser la de seleccionar al personal de campo dentro del área donde se efectúe la encuesta, pero no en los pueblos o lugares donde se vayan a realizar las entrevistas. De este modo, se consigue un personal de campo integrado con el entorno cultural, sin introducir las interferencias que puedan ocasionar sus relaciones sociales. El problema, por fácil que puedan parecer las soluciones, no deja de ser complicado.

El género del entrevistador puede llegar a suponer un problema bastante grave, dependiendo del entorno cultural en que se realice la encuesta. Es habitual considerar que si la encuesta es sobre fertilidad, o temas especialmente personales para las mujeres, será mucho mejor que las entrevistas las realicen entrevistadores del mismo sexo. Sin embargo, cuando se trata de encuestas en países islámicos, por ejemplo, el que una mujer viaje sola, para entrevistar a desconocidos, puede llegar a ser culturalmente inaceptable.

También se requieren cualidades en el entrevistador. Siempre será más conveniente que estos sean honestos, diligentes y con facilidad de comprensión, que no lo opuesto. Pero también se requiere el tener una forma física aceptable y buena salud. Lo más habitual es que las condiciones para efectuar los viajes no

sean las más cómodas, como tampoco lo es el alojamiento o las comidas que puedan conseguir. La combinación de todos estos factores puede hacer más agotador de lo que parece el realizar las entrevistas. Frecuentemente se olvidan en los manuales de advertir que el entrevistar a individuos desconocidos, en lugares con los que no se familiarizado, y donde se es desconocido, puede llegar a ser peligroso. Y esta observación adquiere un relieve especial en los sectores marginados de los países desarrollados, aún cuando frecuentemente se olvide este aspecto.

Es importante controlar la edad de los entrevistadores. En muchas zonas rurales es frecuente encontrar una relación entre status y edad. Si los entrevistados son demasiado jóvenes, pueden encontrar serios problemas para conseguir las entrevistas de los individuos de más edad que ellos. Sin embargo, una vez seleccionado el personal de campo con unos criterios adecuados para el entorno cultural y el tema en estudio, es el momento de afrontar un buen adiestramiento que los acerque, lo más posible, a un entrevistador profesional con el mayor grado de normalización.

La actividad de entrevistar requiere que el entrevistador conjugue tres roles: el de un atento y amistoso “oyente”, el de “profesor”, y el de “profesional de los datos. Se debe ser amistoso para que el rol de “oyente” no pueda ser percibido como amenazante por el entrevistado, pero también debe ser un poco autoritario de modo que pueda conseguir que el entrevistado se ajuste a lo que pide la entrevista en un límite de tiempo razonable. De algún modo, también se debe de efectuar una prudente actividad docente. *“El principal componente de la tarea del entrevistador bajo el rol de profesor, es el de indicarle cuando las respuestas son adecuadas y cuando son incompletas, disgresiones o inapropiadas”* (ISR Newsletter, 1977). Por otra parte, el entrevistador debe de restringir la posible verbosidad del entrevistado, mientras que recoge los datos.

Siguiendo a Singleton (1988), el proceso de formación de los entrevistadores debe de cubrir los siguientes aspectos.

- (a) Proveer a los entrevistadores de información acerca del propósito general del estudio, quién los financia, plan de muestreo y previsiones de uso o publicación.
- (b) Enseñar técnicas y reglas básicas de entrevistar tales como:
 - Habilidad para acceder al entrevistado y obtener su cooperación;
 - Obtener un buen clima en la entrevista sin ser excesivamente amistoso;
 - Preguntar y efectuar “probings” sin que sesgue las respuestas del entrevistado;
 - Saber cómo manejar las interrogaciones y las disgresiones, etc.
- (c) Familiarizar a los entrevistadores con el cuestionario y las instrucciones que existan para su uso, como tarjetas, preguntas, filtros, etc.. Se debe

comentar todo el cuestionario, pregunta a pregunta, dejando bien clara la importancia de lo que se pregunta e indicando cómo anotar respuestas. Pueden producirse posibles respuestas problemáticas y el entrevistador debe de estar prevenido sobre cómo hacerlas frente.

- (d) Hacer demostraciones y supervisar varias prácticas con el cuestionario. Una vez efectuada la explicación del cuestionario, los entrevistadores se aplicarán el cuestionario unos a otros, bajo la supervisión del investigador, con la finalidad de incrementar la familiaridad y experiencia de los entrevistadores.
- (e) Prescindir de aquellos entrevistadores que no posean la motivación y la habilidad para hacer un trabajo aceptable.

El modo de impartir esta formación a los entrevistadores consiste en varias reuniones de grupo más o menos informales. En ellas se efectúan sucesivamente los pasos que se han indicado, explicando la finalidad del estudio, comentando las reglas y técnicas de la entrevista, familiarizando a los entrevistadores con el cuestionario, etc.

Este proceso de adiestramiento permite que los entrevistadores adquieran la capacidad de adoptar un rol profesional en la medida que han adquirido práctica y experiencia. En primer lugar, con el hecho mismo de la investigación social, y en segundo, con el instrumento de recolección de datos que es el cuestionario. Para diseñar la logística del trabajo de campo, hay que tomar una decisión sobre que estrategia puede evitar en mayor grado el generar un mayor error no muestral. Hay que tener mucho cuidado al elegir las fechas de realización del campo. Pueden existir festividades locales que dificulten el campo, tanto en lo que se refiere a la disposición a ser entrevistado de un modo serio, como a problemas de rutas, alojamientos, etc.

Resulta evidente que es necesario y aconsejable efectuar un cierto control de calidad sobre la información que nos entregan los entrevistadores. Esto implica que debe de supervisarse la actividad de los entrevistadores.

En lo que se refiere a la organización del campo, lo primero es elaborar y conservar un archivo con todos los entrevistadores. El mantener un archivo con los entrevistadores es esencial. Este archivo recogerá, para cada entrevistador, el número de cuestionarios recogidos, las entrevistas que ha realizado, la calidad del trabajo, cuánta supervisión han requerido, qué porcentaje de rechazos ha tenido y cuánto se le ha pagado. Este archivo permite el mantener una información básica sobre el estado del personal de campo, necesaria para una buena organización de éste y una óptima asignación de recursos, dando más cuestionarios y entrevistas a los mejores entrevistadores.

En segundo lugar, asegurarse que las entrevistas se han efectuado correctamente, de acuerdo con el cuestionario. Evidentemente, es de sentido común comprobar

que se han efectuado las entrevistas. Para ello, se seleccionará un grupo entre los cuestionarios ya cumplimentados y se contactará con los entrevistados, confirmando que se les hizo la entrevista, y preguntando de nuevo una serie de interrogantes preseleccionadas para corroborar que se ha aplicado todo el cuestionario.

Tercero, mantener contactos regulares con los entrevistadores, recogiendo los cuestionarios ya cumplimentados y comprobando que se han contestado a todas las preguntas, así como que se hayan marcado correcta y claramente las respuestas. Desde una perspectiva ideal, los cuestionarios entregados deben de ser inmediatamente supervisados, en busca de cuestionarios incompletos o mal señalizados. Esto es muy útil al comienzo del trabajo de campo, para que los entrevistadores recojan el “*feedback*” de la importancia que tiene el que los cuestionarios estén cumplimentados completa y correctamente.

Cuando es posible, conviene mantener entrevistas con los entrevistadores para recoger información desde el campo y efectuar un refuerzo psicológico que mantenga en activo el interés de los entrevistadores. Hay que prever el posible sentimiento de soledad del entrevistador y los efectos sobre la entrevista. Paradójicamente, se podría pensar que conforme más experiencia ganan los entrevistadores se produce una mayor calidad de la información. Sin embargo, la tendencia parece ser en sentido contrario. Según los estudios de Cannell (1973), cuantas más entrevistas haya realizado un entrevistador, mayor es la tendencia a descuidar la interacción y la calidad de la información.

Es importante mantener la atención de los entrevistadores con entrevistas que les apoyen moralmente y mantengan un buen nivel de motivación. Por último, permanecer disponible para ser localizado por los entrevistadores, ya sea para responder a posibles dudas o facilitar alguna ayuda que se pueda necesitar. Si la supervisión y control del campo tiene algún sentido (y lo posee), es el de mantener la calidad de la información. En ese sentido, no se trata en absoluto de efectuar un control policial sobre un entrevistador mentiroso, sino de apoyar al equipo de campo, detectando posibles estrategias para mejorar la recolección de la información. Esto último supone, por cierto, el practicar todo el apoyo posible a los entrevistadores.

Una costumbre técnicamente muy provechosa, y en la práctica poco utilizada, es recoger las impresiones del personal del campo acerca de los problemas que pueden haber surgido en la administración del cuestionario. Es evidente que la información acerca de posibles problemas, (como por ejemplo de redacción, insuficiencia de opciones a una respuesta, o de solapamientos entre las categorías de cierre) que se producen especialmente en determinados segmentos de la población, pueden escapar al pretest. Esa información es primordial para futuros estudios, y poder prever la solución más factible, pero también lo es para poder cualificar correctamente la información que hemos recogido con la encuesta. Expresado en otras palabras, la mejor información sobre los datos que se han

obtenido con la encuesta la poseen los entrevistadores. Son ellos quienes pueden aclarar la existencia de respuestas sospechosas.

Ciertamente, los entrevistadores deben de ser motivados para que no presenten reticencias a dar una información que podrían suponer dan mala imagen de su labor como entrevistadores. Para ello, basta insistir en que se han detectado problemas con el cuestionario y es conveniente que ayuden a localizar dónde pudo estar “*la zona de fricción*” entre las preguntas del entrevistador y las respuestas de los entrevistados. Esta práctica, que en general esta poco extendida, sorprende por la riqueza de información que proporciona, así como por las mejoras sensibles que introduce en el ámbito de la redacción y cierre de las preguntas. La pregunta al plantear el diseño era del tipo, “*¿Qué se puede hacer para recoger información válida y fiable?*”, la que ahora corresponde es “*¿Qué información hemos conseguido y que calidad posee?*”.

Es evidente que la posibilidad de recoger la información que poseen los entrevistadores acerca de la entrevista es una potencialidad a aprovechar. Esta opción se convierte en necesidad, dadas determinadas circunstancias. Una encuesta con un alto índice de “*no respuesta*”, requiere, exige, una reunión con los entrevistadores que aclare el porqué de esa situación.

2. LA PREGUNTA

Las posibles influencias que se ejerzan sobre el entrevistado no proceden exclusivamente desde el ámbito del entrevistador; el diseño del cuestionario tiene también un peso importante. Como consideraremos más adelante, el fraseo (wording), o el modo en que se redacte una pregunta, influye en las respuestas que se obtienen, en la medida que el entrevistado puede sentirse presionado a contestar de un modo u otro.

2.1. LA FORMULACIÓN DE LAS PREGUNTAS

Como ya sabemos, la esencia de la entrevista es el juego de preguntas y respuestas entre el entrevistador y el entrevistado. Las preguntas por sí mismas pueden ser una fuente de error, en la medida que desorienten al entrevistado, sean demasiado ambiguas, pidan una información de la que el entrevistado no disponga o no pueda recordar con seguridad, o cuando sean demasiado sensibles y provoquen que el entrevistado las evite o las conteste engañosamente.

La redacción de la pregunta es un tema que requiere de mucho cuidado dada la gran influencia que puede ejercer en las respuestas de los entrevistados. Por ello, en las entrevistas no estructuradas, el hecho de que los entrevistadores puedan cambiar la redacción de las preguntas libremente, cuando piensan que el entrevistado no ha entendido la cuestión, conlleva el que, en esencia, a cada entrevistado se le haga una pregunta diferente. Esto deja un margen muy amplio

a cometer errores al comparar o combinar respuestas que se corresponden formalmente con diferentes preguntas. Pero en este caso, en el que el entrevistado goza de la libertad de reformular las preguntas, no solamente se pueden introducir errores al comparar respuestas a diferentes preguntas, sino que otro motivo de error puede surgir en el momento mismo de redefinir los temas de modo más comprensible. Cuando un entrevistado no entiende algo concreto, el entrevistador intenta definirle o aclararle la cuestión. Si fuese un solo entrevistador el que efectuase las entrevistas, se generaría la misma interpretación para las cuestiones conflictivas, pero es esta una situación muy excepcional. Lo habitual es que diferentes entrevistadores realicen las entrevistas, lo que abre la posibilidad de que los entrevistadores interpreten de forma diferente los temas cuya información deben cubrir. Como ya sabemos, la flexibilidad que se le concede a un entrevistador en las entrevistas no estructuradas facilita la comprensión de los temas por parte de los entrevistados, pero a su vez, deja un margen de incertidumbre con respecto a la comprensión que ellos mismos tengan sobre la información a cubrir en la investigación. Evidentemente, la pregunta cuanto más flexible, más incertidumbre. Una pregunta que se lee plantea el problema de lo que interpreta el entrevistado, mientras que una pregunta que pueda ser aclarada por el entrevistador, presenta el problema por duplicado, al incorporar lo que cree el entrevistador que esta preguntando.

2.2. LA SECUENCIA

Ya en términos de diseño del cuestionario, el orden que ocupe una pregunta en relación a las demás es muy importante, dado el carácter acumulativo de la información que se produce en la entrevista. Una vez enunciada una pregunta, esta será tenida en cuenta por el entrevistado para valorar las siguientes que se le hagan, de modo que influirá en la percepción de su contenido.

En el caso de los cuestionarios autoaplicados, una fuente de error es el hecho de que todas las preguntas pueden ser leídas por el entrevistado antes de comenzar a contestarlas, de manera que desaparece la secuencia que se les puede imprimir mediante una entrevista. Es evidente que en un cuestionario autoaplicado es muy difícil controlar el efecto de una pregunta sobre las demás, y ello debe de tenerse en cuenta al elegir esa técnica.

2.3. ESTRUCTURA

Los errores de estructura aparecen especialmente en los cuestionarios multi-temáticos. La situación más óptima que se puede presentar es la de elaborar un cuestionario para averiguar acerca de un solo tema. Sin embargo, lo más habitual es que, ya sea por motivos económicos o por motivos de urgencia, se plantea la necesidad de conjuntar diferentes temas en un mismo cuestionario. Esto implica que habrá que preguntar sobre temas más o menos dispares. No siempre

se está posibilitado para efectuar transiciones lógicas entre temas, lo que puede provocar incoherencia y confusión en el entrevistado, al tener que contestar, por ejemplo, a preguntas tales como, preferencias políticas, drogas y opinión sobre la reforma fiscal. Este es un problema de difícil solución que debe ser tenido en cuenta. En todo caso, el cuestionario debe contener tanto las preguntas que interesan sustantivamente para la investigación como otras destinadas al control de la calidad de la información obtenida. Para ello, la primera tarea es establecer lo que se denomina una lista temática con las:

1. Preguntas: Especificar claramente qué se quiere saber y las razones para ello (explotación posterior).
2. Variables de control de entrevistado (edad, género, etc.)
3. Variables de control de cuestionario (número de cuestionario, número de estudio, datos entrevistador, etc.)
4. Control de calidad de la entrevista (valoración por parte del entrevistado, duración de ésta, condiciones de realización, rechazos, etc)

2.4. EL ENTREVISTADO

Vamos a suponer que la finalidad de la entrevista ha quedado lo suficientemente clara al entrevistado, de modo que éste colabore sin que se introduzca sesgo en sus respuestas. Estamos en una buena situación para obtener una información válida y fiable. Para ello, el entrevistado debe dar respuestas completas y sinceras. De hecho, un entrevistado puede no dar su opinión sin que por ello exista ningún deseo de engañar o mentir.

Los sesgos que se puedan introducir en la información que se recoja en la encuesta no tienen por qué corresponder con intentos de falsear la información por parte de los entrevistados. Muchas veces, el problema de las respuestas inciertas viene dado por la distorsión que el tiempo haya podido efectuar en la memoria del individuo. Otras veces, es una consecuencia de factores sociales que competen al individuo a un cierto tipo de respuesta. Veamos en detalle algunas de las presentaciones más frecuentes de esto anterior.

2.5. LAS NORMAS SOCIALES

Por ejemplo, en algunos países, la cortesía es un valor muy importante. Esta cortesía implica el evitar mostrarse abiertamente en desacuerdo con el entrevistador, de modo que con la finalidad de mantener una atmósfera agradable, puede tender a adaptar sus respuestas a lo que cree que este desea. Tal como lo sintetiza Mamdami (1972) “*Jamas nadie se marchará de este pueblo diciendo que no somos hospitalarios*”. El deseo de agradar al entrevistador es otra forma de cortesía, donde se busca la aprobación de este. Otra alternativa es el deseo de embromar al entrevistador, sin que haya un deseo real de mentir sino más bien de tomarle el pelo, poniéndolo en situaciones embarazosas.

Estas situaciones anteriores pueden plantearse puntualmente en una encuesta que se efectúe en un país desarrollado. Sin embargo, en determinados segmentos sociales esos tipos de comportamiento pueden ser un rasgo cultural, donde el trato que se le dé al entrevistador, así como la forma habitual de responder a las preguntas, está reglado por las normas sociales vigentes en esas sociedades. Mitchell comenta, por ejemplo, que en el sudeste asiático la gente tiende a infravalorar sus logros, mientras que en otros lugares tienden a sobrevalorar sistemáticamente sus méritos personales. En ese sentido, los entrevistados dan información errónea, pero no por deseo de mentir u ocultar algo, sino simplemente por adecuarse a una norma cultural de la sociedad en la que vive.

2.6. LA FALSEDAD DE LA RESPUESTA

Los errores que se introducen en la respuesta del entrevistado pueden aparecer por dos causas. La primera de ellas cuando el entrevistado es incapaz de dar una respuesta fiable o válida, a pesar de su voluntad; en la segunda, el entrevistado puede desear dar una respuesta equivocada.

a) Incapacidad del entrevistado

La incapacidad del entrevistado para dar una respuesta puede provenir de una posible ignorancia del tema, por olvido o por ser incapaz de articular coherentemente una respuesta.

La ignorancia puede aparecer cuando se interroga al entrevistado sobre cuestiones que desconoce, y acerca de los que darán alguna información exclusivamente para evitar el posible ridículo de parecer desinformado. La posibilidad de que el entrevistado se haya olvidado de aquello por lo que se le preguntó también puede ser alta. Supongamos que se interroga a un entrevistado sobre su recuerdo de voto en todas las elecciones generales que ha participado. En el caso que haya votado a diferentes partidos, lo más probable es que no recuerde a cuáles, ni en qué orden. Idealmente, se debería de preguntar exclusivamente sobre comportamientos recientes. De hecho, cuando se pregunta por comportamientos, se tiende a contestar preferentemente aquellos que se han efectuado últimamente. Así, a una pregunta del estilo, “¿a qué dedica sus momentos de ocio?”, si la actividad más reciente fue a jugar al tenis, esta puede ser la respuesta más probable aún cuando sea un deporte que practique más esporádicamente que otros. Una buena vía de evitar ese problema es preguntar por situaciones específicas, bien localizadas en el tiempo.

Un fenómeno opuesto es el denominado “*telescoping*”, donde el entrevistado da información sobre un fenómeno bastante alejado en el tiempo como si hubiese sucedido recientemente. El efecto de “*telescoping*” alude a la semejanza de aplicar un telescopio sobre el tiempo pasado, con la consecuencia inmediata de aproximar sucesos. Estos se recuerdan, como mucho, más recientes de

lo que realmente son. El investigador, cuando trata con diseños que no son sincrónicos, sino que pretenden recoger la trayectoria de los individuos (por ejemplo, en estudios de población), debe tener muy presente este efecto, en la medida que puede romper la secuencia de sucesos que, desde el punto de vista del análisis, tiene sentido teórico. Muy frecuentemente, es más fácil recordar un hecho o una impresión, que el situarlo en el tiempo. Este error es difícil de detectar y afecta a la regularidad con que, según el entrevistado, puede producirse un comportamiento.

Otro posible fallo de la memoria es el de promediación (averaging). La tendencia de promediar los recuerdos está bastante extendida. En ella, se responde a una pregunta concreta de acuerdo al comportamiento que ha sido más usual en el entrevistado. Si frecuentemente va los domingos a misa, y el último domingo hubo elecciones parlamentarias, la pregunta sobre qué hizo el domingo pasado antes de ir a votar puede tener como respuesta la de ir a misa, aún cuando no fuese cierto, porque ese domingo no fue. El fenómeno de no responder concretamente a preguntas concretas, sino aplicar lo que es una práctica usual en el entrevistado, debe de vigilarse para evitar el llegar a conclusiones erróneas al asociar sucesos puntuales que no se produjeron conjuntamente.

Por último, en lo que se refiere a los problemas con la memoria, se puede producir un problema de omisión. La omisión aparece cuando no se recuerda, y por lo tanto, no se indica parte de la información que es relevante en relación al tema objeto de la pregunta. La omisión es una circunstancia que puede presentarse con un carácter más general que el de un mero olvido, en la medida que podría ser, a sí mismo, una selección consciente de la información que el entrevistado considera realmente relevante y que no tiene porque coincidir con aquello que lo es para el investigador.

Tal y como indicábamos, la tercera de las situaciones por las que el entrevistado puede dar una información equivocada es debido a la incapacidad de expresar la respuesta correcta. Esto sucede frecuentemente cuando se trata de preguntar por las motivaciones que existen para un determinado comportamiento. Evidentemente, la incidencia es mucho más alta cuando la pregunta es abierta, y el entrevistado no encuentra la fórmula correcta para expresar algo que no acaba de verbalizar.

La intención es conseguir siempre respuestas completas y razonadas, pero esto no se produce espontáneamente y en algunas situaciones se recoge solamente información superficial, especialmente a causa de la incidencia de este último fenómeno que hemos descrito.

b) Falsedad de la respuesta

En determinadas circunstancias, el entrevistado da respuestas que no se ajustan a la verdad de aquello que piensa realmente. Esto se presenta corrientemen-

te cuando los cuestionarios son largos y aburridos, de modo que la entrevista se convierte en algo enojoso para el entrevistado; incomodar al entrevistado es una forma de provocar una serie de errores que pueden afectar sensiblemente a la fiabilidad y validez de la información que se obtiene.

Los motivos por los que el entrevistado puede contestar falsamente son bastante variados. Así, un cuestionario excesivamente largo y aburrido puede indisponer al entrevistado, de forma que su colaboración se vea condicionada desfavorablemente. De hecho, algunos de los aspectos que podían influir para no aceptar el ser entrevistado son de por sí factores que pueden introducir sesgos en las respuestas en el transcurso de las entrevistas. Por ejemplo, el disgusto a ser molestado y la invasión de la vida privada que podía haber provocado el rechazo de la entrevista, cuando ésta ya se está desarrollando incidirá en un posible falseamiento de las respuestas. Generalmente a los individuos no les preocupa el ser entrevistados, pero podrían molestarse cuando las preguntas se dirigen hacia temas personales. Aproximadamente el 20% de los entrevistados, mediante entrevistas personales (“cara a cara” y telefónicas), se niegan a contestar preguntas tales como los ingresos que obtienen. Otros simplemente distorsionaran las respuestas. Parece lógico que determinados temas no se consideren confiables a un desconocido, por más que se trate de una entrevista para una investigación con finalidad académica.

Otro motivo que puede producir respuestas falsas es el de la fatiga. Conforme la entrevista se prolonga es más fácil que la fiabilidad de las respuestas decline. Especialmente para aquellos entrevistados que en la etapa de motivación mostraban reticencias a ser entrevistados por falta de tiempo o cansancio. Estos entrevistados tenderán a dar respuestas cada vez más abruptas y secas, con la intención de finalizar la entrevista en el plazo más breve posible. La duración de una entrevista es algo más bien subjetivo, y mientras que si el tema de la entrevista es divertido e interesante una hora puede transcurrir muy de prisa, de ser un tema aburrido, un período de 15 minutos daría la impresión de ser muy prolongado. Los errores que se introduzcan podrían compensarse de producirse aleatoriamente; no ocurriría esto de ser sistemáticos, dándose una tendencia a concentrarse en algún punto concreto del cuestionario.

El entrevistado también puede sentirse forzado a mentir en algunas ocasiones, ya sea ajustando su respuesta de modo que refleje la deseabilidad social, o intentando elevar su prestigio en la opinión del entrevistador.

Por ejemplo, en los hábitos de contenido cultural, como lectura, el entrevistado puede mentir con la finalidad de aparentar aquello que cree es lo más deseable socialmente. En un sentido opuesto, el entrevistado intentará evitar mencionar aquellos aspectos que se consideren socialmente incorrectos, tales como consumo de bebidas alcohólicas.

Como ya sabemos, el entrevistado puede mentir por cortesía, expresando opi-

niones exclusivamente porque éstas puedan parecer de interés para el entrevistador. Del mismo modo, el hecho mismo de formular una pregunta parece exigir la existencia de la respuesta, por lo que el entrevistado puede sentirse obligado a contestar incluso cuando desconozca sobre qué le están preguntando.

Por último, otra vía de producir respuestas falsas es la denominada estilos de respuesta. Esta aparece cuando se muestra por parte del entrevistado una predisposición a falsear determinadas categorías de respuesta, independientemente del contenido de lo que se pregunta. En general, los cierres del tipo “*de acuerdo*”, “*desacuerdo*”, o “*si*” “*no*”, son los que presentan una predisposición a este tipo de sesgo.

2.7. LA NO RESPUESTA

La no respuesta gravita como una posible fuente de error en cualquier encuesta que se efectúe. Las causas que pueden influir para rehusar a ser entrevistado o no contestar determinadas preguntas son muy variadas, así como las posiciones metodológicas que se puede adoptar ante ella.

La no respuesta procede de dos fuentes diferentes. Una de ellas es conocida como “*fenotípica*”, según la cual los rechazos se corresponden con las características del procedimiento de recolección de datos (qué se pregunta, cómo se pregunta, duración de la entrevista, etc). La incidencia de este tipo de no respuesta variará de estudio en estudio, pues los segmentos de población que no responden son diferentes.

La otra fuente de no respuesta es de carácter “*genotípico*”. Se refiere a las características endógenas de los individuos (edad, sexo, ocupación, etc.) que establemente rechazan ser entrevistados. Este tipo de no respuesta permanece constante a lo largo de diferentes estudios.

2.8. DEPENDENCIA E INDEPENDENCIA

Un aspecto de importancia primordial es la relación que pudiera existir entre los segmentos de individuos que no responden y el tema objeto de investigación. Así, por ejemplo, pudiera suceder que (en el caso de la “*no respuesta*” genérica) rechazan ser entrevistados los que nos interesan estudiar. Si resulta que la población que se quiere estudiar se niega a ser entrevistada, difícilmente obtendremos alguna información fiable o válida sobre ellos. Tal como indicábamos en páginas anteriores, la encuesta es particularmente apropiada para las clases medias en las sociedades desarrolladas. El plantear investigaciones exige que conjuntamente se intente detectar si la población a estudiar es receptiva al método empleado para ello. Una primera pregunta es si aquellos que se desean investigar están realmente entre los cuestionarios recogidos o si desaparecieron mediante el rechazo a ser entrevistados.

Otro problema importante se produce cuando el tema, o los temas tratados,

son de los considerados sensibles. Esto puede plantearse incluso en preguntas muy concretas.

Es evidente que la posibilidad de recoger la información que acerca de la entrevista poseen los entrevistadores es una potencialidad a aprovechar. Esta opción se convierte en necesidad dadas determinadas circunstancias. Una encuesta con un alto índice de “no respuesta”, requiere, exige, una reunión con los entrevistadores que aclare el porqué de esa situación, ya sea para cualificar los segmentos de individuos reacios a contestar, o detectando los problemas, si los hubiese, que produzcan la redacción de las preguntas o las materias tratadas por la encuesta. No solo es una buena estrategia el recoger información de los entrevistadores una vez terminado el trabajo de campo. Aun cuando este se encuentre en vías de realización, el efectuar reuniones periódicas con los entrevistadores aporta una buena visión sobre cómo progresa la encuesta. Ello tanto para detectar problemas de cuestionario o campo, como para comprobar si este es un método de recolección de datos válido para esa sociedad o el tema tratado.

2.9. EL CUESTIONARIO

El instrumento para la recogida estandarizada de datos, tanto por correo como por teléfono o “*cara a cara*”, es el llamado cuestionario. Este consiste en un formulario que contiene escritas una serie de preguntas o afirmaciones, y sobre el que se consignan las respuestas. La redacción y estructuración de las preguntas en el cuestionario es una tarea cuidadosa, pues de ella depende, primero, que se obtenga la información que se desea y no otra y, segundo, no influir en las respuestas de los individuos, ya directa o indirectamente. Los datos deben representar aquello que se pretende y ser obtenidos espontáneamente sin influir en el entrevistado. En ese sentido, la diferencia básica entre una recogida de datos con cuestionario y otra sin e, consiste en que sin cuestionario “*se pregunta*”, mientras que con el cuestionario “*se lee la pregunta*”. Esto ofrece un mayor control sobre lo que se pregunta empleando el cuestionario que haciéndolo directamente. Por este motivo es por el que la recogida de datos con cuestionario se denomina normalizada, en la medida que en todas las administraciones del cuestionario se “*lee*” la misma pregunta.

Este deseo de normalizar la interacción y el intercambio de información puede ser esencial para el buen diseño de cuestionario. Ello significa cuidar la redacción de las preguntas, así como procurar la mayor exhaustividad y diferencias en las respuestas, atender al orden lógico de la presentación de temas en el cuestionario, o a la secuencia con que aparecen las preguntas. Consideremos esto más en detalle.

2.10. DISEÑO DEL CUESTIONARIO

En general, existen una serie de normas que deben tenerse en cuenta en el

momento de plantearse el diseño de un cuestionario. La mayoría de ellas han sido aportadas por la experiencia de los investigadores. Las que siguen son un resumen de las más importantes.

La clave es tener muy claro el propósito del estudio. Esta observación no es en absoluto trivial, dado que garantiza que se va a preguntar aquello que es relevante para la investigación, y evitar el introducir cuestiones que son ajenas o secundarias a lo que realmente importa. En relación a ello, las tareas son las siguientes.

2.11. ESTRUCTURA

En primer lugar, construir una lista con los temas sobre los que se quiere obtener información, sin preocuparse de la redacción o del orden lógico que vayan a seguir las preguntas. A continuación, es preciso numerar los temas anteriores siguiendo un orden lógico. Cada tema será objeto de una o más preguntas. Los criterios más útiles para ordenar la secuencia de preguntas son los siguientes.

- a Hay que conseguir que el entrevistado se sienta motivado para contestar todo el cuestionario. Para ello, un cuestionario se abre siempre con preguntas fáciles de contestar, poco controvertidas o atrayentes, dejando las más conflictivas para secciones más avanzadas del cuestionario.
- b Las preguntas deben sucederse de forma que tengan sentido para el entrevistado. Cada pregunta debe correlacionarse lógicamente a la anterior.
- c La sucesión de las preguntas en el cuestionario deben de ser fáciles de seguir por parte del entrevistador (entrevistado, en el caso que fuese auto-aplicado). Así, las transferencias entre preguntas se deben presentar de un modo claro, de forma que el entrevistador no se pierda entre las preguntas que debe o no efectuar.

Si alguna pregunta puede influir en las respuestas que se den a otras preguntas, es conveniente situar la que influya detrás de las influenciadas.

1. Las cuestiones que tengan mayor probabilidad de no ser contestadas, que puedan ser contestadas erróneamente (como nivel de ingresos familiar), o que puedan poner al entrevistado a la defensiva, deben situarse al final.
2. En general, las preguntas que tengan menor importancia deben ir al final. Así, de no ser contestadas por que el entrevistado, se canse o se niegue a seguir, la pérdida de información será menor.
3. Para los cuestionarios excesivamente largos es conveniente, conjuntamente con los aspectos anteriores, evitar los cierres de preguntas que faciliten el responder mecánicamente. Por ejemplo, del tipo “sí-no”, “de acuerdo-desacuerdo”, allí donde ya se sospeche cansancio o aburrimiento por parte del entrevistado.

Una vez definida la estructura del cuestionario se puede comenzar ya con la redacción de las preguntas. Al igual que hicimos con la estructura de las preguntas, vamos a dar una serie de orientaciones sobre la redacción de éstas.

2.12. LA REDACCIÓN DE LAS PREGUNTAS

La redacción de las preguntas es uno de los aspectos más esenciales del diseño del cuestionario, en tanto que supone la auténtica línea de contacto con la realidad social, cuando la recolección de datos se basa en información transmitida ya sea oralmente o por escrito. Durante mucho tiempo, la formulación de las preguntas se ha considerado como una técnica a medio camino entre el arte y la ciencia. Valga como ejemplo *“The art of asking Questions”* (1951), de Payne, donde, a pesar del título, se presenta una buena combinación entre arte y ciencia. En la actualidad, podemos considerar que la investigación sistemática que se lleva a cabo sobre los efectos de la redacción de las preguntas en los entrevistados, así como la ya considerable experiencia en este campo, que evita el introducir errores en muchas situaciones, sugieren más su consideración como técnica que como arte.

Existen muchas sugerencias para evitar los errores más graves que se pueden producir al redactar un cuestionario. Así, Warwick y Lininger (1975) sugieren las siguientes ocho interrogantes que debe efectuarse el investigador sobre las preguntas y su estructura.

- ¿Son las palabras utilizadas en la pregunta, simples, directas y familiares a todos los entrevistados?
- ¿Son las preguntas tan claras y específicas como es posible?
- ¿Intentan los ítems cubrir más de un punto dentro de una misma pregunta?
- ¿Son alguna de las preguntas tendenciosas o con dobles sentidos? ¿Emplean palabras cargadas emocionalmente o que amenacen la autoestima?
- ¿Es la pregunta aplicable a los entrevistados a los que se le preguntará?
- ¿Contestarán los entrevistados la pregunta de un modo sesgado, tal como decir “sí”, independientemente de su contenido, o darán respuestas que son socialmente aceptables más que los propios puntos de vista de los entrevistados?
- ¿Pueden acortarse las preguntas sin que haya pérdida de significado?
- ¿Se leen bien las preguntas?

En general, las ideas esenciales que deben tenerse presentes al diseñar un cuestionario son las siguientes.

- Cada pregunta debe plantear un solo tema.
- Las preguntas deben ser claras, simples y concisas. Es conveniente evitar las preguntas demasiado largas.
- Algunas veces hay que decidir sobre preguntas abiertas o cerradas, sobre

todo cuando la información es bastante incompleta. En el caso de optar por preguntas abiertas, el “*wording*” debe de ser aún más cuidadoso de lo habitual, para evitar diferentes interpretaciones de la pregunta, lo que produciría respuestas incomparables (con poca lógica de codificación) y su consiguiente anulación.

- Cuando se formula una pregunta cerrada, las respuestas ofrecidas deben ser las más apropiadas para la pregunta, intentando cubrir todo el rango de respuestas que el entrevistado pueda elegir.

Hay que evitar con especial cuidado que las respuestas puedan aparecer duplicadas, o no lo suficientemente excluyentes entre sí.

- Las palabras utilizadas deben tener el mismo significado para todos los entrevistados.
- El vocabulario empleado debe ser el adecuado a la cultura y status educativo de los entrevistados. Por ejemplo, si los entrevistados son médicos, deben emplearse términos médicos.
- Las preguntas deben evitar ser tendenciosas. Hay que considerar que diferentes redacciones de una pregunta produce diferentes respuestas.
- Las preguntas deben proteger el “*ego*” del entrevistado, evitando mediante la redacción el que se sienta agredido. Soluciones normales son del tipo “*Mucha gente...*”.

Estos son los aspectos a cuidar en lo que a redacción se refiere. En todo caso, las observaciones generales deben prever la posibilidad de que las aplicaciones técnicas no son directamente exportables. Así, Bulner (1982) comenta una situación bastante interesante, consecuencia de la aplicación indiscriminada de observaciones técnicas netamente occidentales. En una encuesta nacional en Perú, los investigadores incluyeron una versión traducida de la “*ladder scale of aspirations*” de Cantril. En esta técnica, después de mostrar al entrevistado el dibujo de una escalera, se le indica que la posición más alta es la mejor vida que él puede imaginar, mientras que la posición más baja es la peor vida posible. Tras esta explicación del instrumento, se preguntan tres aspectos: ¿dónde se encuentra en la actualidad?, ¿dónde estaba hace cinco años?, ¿dónde estará en cinco años?².

2. En cierto modo, se trata de una aplicación explícita de las observaciones de Montesquieu en su “*Grandeza y decadencia del pueblo romano*”, donde para ejemplificar la percepción de decadencia social planteaba que “nuestros antepasados eran mejores que nuestros padres, y estos mejores que nosotros. Seguramente, nuestros hijos serán peores”. Cantril formaliza dicha observación haciéndola operativa para la investigación social.

Con esta técnica de medición surgieron varios problemas. Muchos entrevistados miraron el dibujo, pero no vieron en él ninguna escalera. Otros, tal como ya sabemos con respecto a las formulaciones condicionales, entendieron la tarea, pero no encontraron fácil el planteamiento hipotético de las preguntas. Un campesino del sudeste del Perú, que pareció entender la pregunta, señaló uno de los peldaños más altos para ayudar a entender su opinión. Sin embargo, cuando se le preguntó dónde estaba hace 5 años, señaló el horizonte y contestó “*en Ayacucho*”. Un problema parecido mostraban los ancianos al preguntarles dónde estarían dentro de 5 años. Algunos señalando hacia arriba contestaron “*en el cielo*”.

Otro de los asuntos a decidir es el grado de libertad que se va a dejar al entrevistado para contestar con sus propias palabras o no.

2.13. LAS PREGUNTAS ABIERTAS Y CERRADAS

Como ya hemos considerado, en el método de la encuesta se produce una interacción entre el investigador y el entrevistado, normalmente basada en una serie de preguntas y respuestas. Tanto las preguntas como las respuestas pueden estar sujetas a un grado diverso de estructuración y formalización.

En el procedimiento de la entrevista abierta, las preguntas pueden ser reformuladas o introducidas en momentos diferentes, y las respuestas son extensas incorporando las palabras, argumentos o reflexiones del propio entrevistado. Como es fácil de suponer, este sistema de recolección de la información repercute en una menor comparabilidad entre las respuestas de los sujetos, estadísticamente hablando, al complicar la creación de categorías y codificación de la información. Asimismo, repercute en una mayor dificultad de análisis, en la medida que las reformulaciones de las preguntas impliquen matices que introduzcan nuevos aspectos, o en el peor de los casos, la cuestión sea interpretada de modo diferente.

La formulación más rígida es aquella que establece el empleo de cuestionario, implicando con ello la estricta lectura del enunciado de la pregunta, en el orden que vienen establecidas, así como el ofrecimiento de respuestas alternativas entre las que deberá escoger el entrevistado. En el caso de que un entrevistado no entiende una pregunta, el entrevistador sólo puede limitarse a repetirla, quizás más despacio o pronunciando más claro, pero sin introducir nuevas redacciones. Este procedimiento mejora la comparabilidad entre sujetos y facilita el análisis, pero exige un buen conocimiento de las posibles respuestas que pueda ofrecer el entrevistado.

Entre los dos procedimientos extremos que hemos introducido, existe un rango de posibilidades que incorpora aspectos de ambos. Una entrevista sin cuestionario puede estar más o menos estructurada, o en un cuestionario aparecer preguntas abiertas para ser contestadas por el entrevistado. El empleo de una

estrategia u otra, de un juego de preguntas y respuestas, así como la posible combinación de ambas dependerá del tipo de información que se pretende obtener en cada momento.

Las preguntas en una investigación son, generalmente, abiertas o cerradas, según se ofrezca o no una alternativa de respuesta, aún cuando también existe la posibilidad de dejar una pregunta semiabierta. Por ejemplo, enumerando una serie de respuestas que se sabe insuficientes y agregando la opción “*otra*” (especificar), con lo que la pregunta queda semiabierta. En una pregunta cerrada, se presentan una serie de respuestas entre las que el entrevistado deberá elegir la que considere que corresponde mejor con lo que piensa u opina. Una pregunta abierta deja la respuesta, cualquiera que esta sea, en manos del entrevistado. La formulación de una pregunta abierta o cerrada está sujeta a la evaluación de las ventajas o inconvenientes que puedan presentar en cada momento concreto. Con carácter general Nachmias (1976) destaca las siguientes.

En el caso de las preguntas cerradas, el hecho de que estas son fáciles de preguntar y rápidas de contestar, evitando que se pierda tiempo al escribir las respuestas, es una ventaja importante. Su interpretación, como ya hemos indicado, es rápida y con pocas complicaciones. Entre las desventajas más destacables se encuentra, primero, la de forzar la elección entre opciones ya dadas y, segundo, que puede sugerir al entrevistado posibles respuestas en las que nunca habría pensado.

Por ello, una de las ventajas de las preguntas abiertas es la posibilidad que tiene el entrevistado de contestar libremente, de modo espontáneo, y según sus propias palabras. Son mucho más flexibles, de forma que en el caso que la respuesta sea poco clara, es posible pedir al entrevistado que la aclare o se extienda un poco más en ella. El plantear preguntas abiertas acerca más al entrevistador con el entrevistado, lo que favorece el completar todas las preguntas así como que sean contestadas con mayor veracidad. A esta serie de ventajas se unen algunas desventajas. La primera de ellas, que son difíciles de contestar, y aún más difíciles de analizar. Es preciso crear una codificación para clasificar las diferentes respuestas, en cuyo proceso es preciso prescindir de los detalles y matices de la información que facilitó el entrevistado. Incluso puede darse alguna confusión sobre lo que el entrevistado quiso realmente decir.

El grado en que el cuestionario esté estructurado es una cuestión importante en las encuestas efectuadas en zonas rurales de los países en vías de desarrollo. Para los entrevistados poco habituados a condensar en categorías concretas sus pensamientos, el empleo de preguntas cerradas puede producir problemas a la hora de obtener respuestas. Sin embargo, la posible solución de utilizar preguntas abiertas hace más difícil el control de las respuestas, a menos que el personal de campo esté muy bien adiestrado y haya asumido perfectamente los objetivos de la investigación. Warwick y Lininger (1975) ofrecen una solución

mixta que combina un principio de cuestionario con preguntas abiertas, introduciendo progresivamente las preguntas cerradas. Finalmente Bulner (1982) observa que *“la elección final entre preguntas abiertas o cerradas depende del tema de estudio de la encuesta, las posibles actitudes de los entrevistados hacia determinadas cuestiones y otras características de los entrevistados, tales como educación”*.

Como ya indicamos, son las necesidades de la investigación las que decidirán el uso de un tipo u otro de pregunta. En general, las consideraciones a tener en cuenta son las ya clásicas propuestas por Lazarsfeld (1944). En primer lugar, depende de los objetivos de la entrevista. Las preguntas cerradas son preferibles cuando el objetivo del investigador es obtener la opinión (acuerdo o desacuerdo) con un cierto punto de vista. La pregunta abierta es más útil cuando el investigador busca comprender el proceso por el que el entrevistado llega a un punto de vista u opinión.

En segundo lugar, la cantidad de información que el investigador tenga sobre el tema en cuestión. Una pregunta abierta, tiene la posibilidad de ofrecer información que es desconocida para el investigador, mientras que una cerrada no.

Tercero, la posibilidad de que el entrevistado posea o no una opinión definida sobre el tema objeto de la pregunta. Si se considera que el entrevistado no posee una opinión clara sobre el tema de interés, por ejemplo, porque este sea muy reciente, es preferible una pregunta abierta, dado que el ofrecer una pregunta cerrada puede inducir a elegir una opción cualquiera, mientras que quizás el pensar sobre ella llevaría a una opinión diferente.

Finalmente, hay que considerar la facilidad de comunicación que posea el contenido de la respuesta, así como el grado de motivación del entrevistado para responder. Las preguntas cerradas no requieren de mucha motivación por parte del entrevistado para ser contestadas y le exponen menos emocionalmente cuando la pregunta es conflictiva. En general, las preguntas cerradas son menos *“agresivas”* que las preguntas abiertas. En los casos que hemos mencionado, baja motivación o temas conflictivos, las preguntas cerradas obtienen una mayor proporción de respuestas que las abiertas.

Preguntas y respuestas son la base de la obtención de información mediante encuesta. Sin embargo, estas no se presentan aisladas, como es de suponer, sino que se ordenan de forma que se suceden unas a otras. Es en la recolección de los datos mediante cuestionario que se produce la mayor estructuración de las preguntas y respuestas. La estructura del cuestionario y el orden en que aparecen las preguntas es una cuestión importante a cuidar, para garantizar al máximo la independencia entre preguntas.

En los asuntos importantes no recogía los votos según el orden establecido,

sino según su gusto, de suerte que cada senador tenía que estar dispuesto a emitir parecer en vez de limitarse a seguir el de otro.

Suetonio. Los doce césares

Además de las orientaciones habituales en el diseño de cuestionarios que hemos considerado, se han definido una serie de reglas útiles para aquellos cuestionarios aplicados en los países en vías de desarrollo. Estas han sido sugeridas por Brislin, Lonner y Thorndike, basándose en estudios psicológicos cross-culturales.

- a) Emplear las frases más cortas y simples posibles, con la longitud siempre menor a 16 palabras.
- b) Utilizar voces activas en lugar de las construcciones pasivas, limitando lo más posible éstas últimas.
- c) Repetir los nombres en lugar de sustituirlos por pronombres.
- d) Evitar metáforas o frases coloquiales. Es difícil que los equivalentes en la sociedad en estudio posean el mismo significado.
- e) Evitar los modos subjuntivos.
- f) Evitar los adverbios y las preposiciones de lugar.
- g) Evitar las formas posesivas, donde sea posible.
- h) Emplear términos específicos en lugar de los genéricos.
- i) Evitar palabras que expresen vaguedad en relación con algún evento, tales como probablemente o frecuentemente.
- j) Evitar frases con dos verbos distintos, especialmente si los verbos sugieren diferentes acciones.

Estas reglas anteriores son bastante lógicas en su aplicación cross-cultural, si bien implican un mayor grado de dificultad y laboriosidad en el diseño del cuestionario. Asimismo, también es lógica la necesidad de reducir el uso de tarjetas, ante la mayor probabilidad de encontrar entrevistados analfabetos.

2.14. LOS FORMATOS DE CIERRE

En la construcción de los cierres de las preguntas, debemos considerar que estos suponen una estimación sobre la respuesta esperada desde el entrevistado. En términos generales, podemos considerar los siguientes formatos de cierre.

- a. Dicotómicas (sí, no; de acuerdo y en desacuerdo con frases)
- b. Ordinales (mucho, bastante, poco, nada; Excelente, bien, regular, mal; Muy satisfecho, bastante satisfecho, algo satisfecho, nada satisfecho).
- c. Formato gráfico (escalas de Cantril, ubicación ideológica, etc.)
- d. Rangos (ordenar las opciones ____ Sabor, ____ Coste, ____ Natural, etc.)
- e. Lista de categorías
- f. Anotación directa (Ejemplo: edad)
- g. Pregunta abierta

Ejemplos de cierres

a) Pregunta con formato de cierre dicotómico

¿Sabe Ud. que en un futuro próximo a través de su teléfono móvil podrá conectarse a Internet?

- Sí 1
- No 2
- N.S. 8
- N.C. 9

b) Pregunta con formato de cierre ordinal

¿En qué medida considera Ud. necesario que cada hogar español disponga de un ordenador?

- Muy necesario 1
- Bastante necesario 2
- Poco necesario 3
- Nada necesario 4
- N.C. 9

c) Pregunta con formato de cierre gráfico

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*).

Izda.		Dcha.										
	01	02	03	04	05	06	07	08	09	10	NS 98	NC 99

d) Establecer rangos

Ordeneme según su importancia las características que según usted, debe poseer un político. (*MOSTRAR TARJETA H*)

- _____ - Honestidad
- _____ - Simpatía
- _____ - Eficacia para resolver problemas
- _____ - Pensar como yo (ideología)
- _____ - Tener ideas nuevas
- _____ - Tener experiencia

e) Lista de categorías con cierre abierto

¿En cuál de las siguientes situaciones se encuentra Ud. actualmente? (MOSTRAR TARJETA E).

- Trabaja..... 1
- Jubilado o pensionista (anteriormente ha trabajado) 2
- Pensionista (anteriormente no ha trabajado, sus labores, etc.) 3
- Parado y ha trabajado antes 4
- Parado y busca su primer empleo 5
- Estudiante 6
- Sus labores 7
- Otra situación, ¿cuál? _____ 8
- _____ 8
- N.C. 9

f) Anotación directa

En una escala de 0 a 10, en la que en el 0 se colocarían los que lo rechazan por completo y en el 10 los absolutamente apasionados, ¿en qué posición se colocaría Ud.? (MOSTRAR TARJETA B).

Posición escala: _____
 N.S. 98
 N.C. 99

g) Pregunta abierta de respuesta múltiple

Para empezar, me gustaría saber ¿cuáles son, a su juicio, los tres problemas principales que existen actualmente en España? (MÁXIMO TRES RESPUESTAS).

N.S. 98
 N.C. 99

Por último, es fundamental que el cuestionario sea probado en el banco de pruebas del pretest, con entrevistados que posean las mismas características que la población a la que se pretende aplicar el cuestionario. Hay que animar a estos entrevistados en el pretest para que expresen todas sus dudas o problemas al contestar el cuestionario.

El pretest mostrará seguramente que es necesario reescribir algunas preguntas, cambiar la estructura del cuestionario, cerrar preguntas que estaban abiertas y eliminar aquellas que no funcionan o que se han mostrado como innecesarias.

En general, el pretest mejora notablemente los cuestionarios al ajustarlos a la realidad social sobre la que se aplicará. Esto que es un principio de general aplicación, es particularmente importante cuando los temas tratados son sensibles o la población objeto no es clase media, donde son esenciales todos los procedimientos que conduzca a un mejor ajuste de la investigación con la realidad social que estudia.

BIBLIOGRAFÍA

- BULMER, M. (1982) *The uses of social research*. Allen and Unwin, London.
- GROVES, ROBERT M. (1989) *Survey Errors and Survey Costs*. Wiley.
- KAHN, ROBERT L., y CHARLES F. CANNELL. (1957) *The Dynamics of Interviewing: Theory, Technique, and Cases*. Wiley.
- HURS-CESAR y ROY, (1976) *Third world surveys: Survey Research in developing nation*. Macmillan of India. Delhi.
- HYMAN, HERBERT H., with WILLIAM, J. COBB, JACOB J. FELDMAN, CLYDE W. HART, y CHARLES HERBERT STEMBER, (1954) *Interviewing in Social Research*. University of Chicago Press.
- LAZARSELD P.F. "The controversy over detailed interviews" *Public Opinion Quarterly*, 8: 38-60
- OPPENHEIM, A.N. (1966) *Questionnaire design and attitude measurement*. Heinemann, London.
- PAYNE, STANLEY L. (1951) *The Art of Asking Questions*. Princeton University Press.
- RALIS, SUCHMAN y GOLDSSEN (1958) "Applicability of survey techniques in Northern India". *Public Opinion Quarterly*, 22. 245-250.
- SCHUMAN, HOWARD, and Stanley Presser. (1981) *Questions and Answers in Attitude Surveys*. Academic Press.
- STEPHAN, FREDERICK F., y PHILIP McCARTHY. (1958) *Sampling Opinions*. Wiley.
- SUDMAN, SEYMOUR, y NORMAN M. BRADBURN. (1974) *Response Effects in Surveys*. Aldine,
- WARWICK, D.P. And LININGER, C. (1975) *The sample survey: Theory and practice*. McGraw-Hill, New York.

TEMA 4

ESCALAMIENTO: MEDIDA DE LAS ACTITUDES, OPINIONES, SENTIMIENTOS Y PERCEPCIONES

Juan Luis Castejón

1. INTRODUCCIÓN

La actitud es un constructo teórico destinado a definir las relaciones entre el sujeto y el objeto. Una característica fundamental de este rasgo, como de otros constructos psicológicos, es que no se pueden medir directamente, sino que se infieren de la conducta, o de las declaraciones verbales del sujeto.

Las actitudes conciernen a los sentimientos hacia ciertos objetos sociales, objetos físicos, personas determinadas, instituciones, etc. Constituyen una mezcla de pensamientos y sentimientos, opiniones o percepciones, acerca de personas u objetos.

Una de las controversias clásicas en la definición de actitud alude a la naturaleza estructural de las actitudes. Existen dos escuelas de pensamiento respecto a la estructura de las actitudes. La escuela del componente único mantiene que una actitud es simplemente la tendencia a evaluar un objeto o constructo en términos positivos o negativos. Una definición representativa de esta tendencia es la de Thurstone (1946), que sugiere que “una actitud es la intensidad del afecto positivo o negativo hacia un objeto psicológico”. Un objeto psicológico es cualquier “símbolo, persona, frase, eslogan o idea hacia el que las personas pueden diferir respecto al afecto positivo o negativo”. Las teorías que conciben las actitudes de esta manera se conocen como unidimensionales, debido a que su concepción de la estructura de la actitud queda restringida a una única dimensión: la evaluativa.

La escuela de los componentes múltiples, conceptualiza la estructura de la actitud, formada por 2 componentes:

- A. Cognoscitivo. Consiste en las percepciones y creencias hacia un objeto. La información que tenemos sobre un objeto constituye el componente cognoscitivo de la actitud hacia él.
- B. Afectivo. El sentimiento en favor o en contra de un objeto. Es el componente más característico de las actitudes. En esto se diferencian de las creencias y opiniones que se caracterizan sobre todo por su componente cognoscitivo, pero no necesariamente por su componente afectivo.

2. ENFOQUES EN LA MEDICIÓN DE LAS ACTITUDES

Conforme se desarrolla el constructo de la actitud, surge la necesidad de desarrollar técnicas válidas para la medición de las actitudes, sobre todo del componente evaluativo.

El método más directo de medir las actitudes es preguntar a las personas de una u otra manera cuáles sus percepciones, sentimientos u opiniones acerca de un determinado objeto de actitud. Ahora bien, considerar una actitud como un constructo, requiere definir la existencia de múltiples indicadores del mismo. Por ejemplo, se les presenta a los sujetos una lista que contiene enunciados favorables y desfavorables con respecto a las Naciones Unidas y se les pide que expresen su grado de acuerdo o desacuerdo con cada uno de ellos. Este conjunto o inventario de enunciados se denominan escalas de evaluación de actitudes.

Se han desarrollado tres grandes métodos para medir actitudes, opiniones, percepciones o sentimientos: los de Thurstone, Likert y Guttman.

De forma resumida, los principios y asunciones de los que parte *Thurstone* para fundamentar su técnica, son los siguientes:

- a. Dado un conjunto de estímulos es posible ordenarlos en un continuo psicológico, en función del grado de atributo que estos estímulos poseen. Según la “ley del juicio comparativo”.
- b. Es posible establecer una escala diferencial o escala de “intervalos aparentemente iguales”, en la que los ítems o enunciados son seleccionados de modo tal que puedan ser escalonados, expresando un continuo psicológico subyacente.

Los sistemas de medida que se utilizan en el *escalamiento* pueden centrarse sobre los sujetos, en los objetos o estímulos, y sobre sujetos y estímulos conjuntamente. La técnica Thurstone utiliza un sistema de medida que *se basa en los objetos*, o ítems, es decir, ordena y mide los distintos estímulos empleados para el estudio de la actitud. Se pide a varios sujetos que emitan juicios con el fin de obtener valores de escala para objetos externos al propio sujeto. El sujeto se considera como observador de objetos (opiniones, enunciados, etc) y se le pide su opinión en orden a posicionar dichos objetos en una escala o dimensión determinada.

El fundamento lógico de la escala de Thurstone lo constituye el modelo probabilístico no monótono, según el cual cada ítem está concebido para representar en un sentido estadístico un punto de un continuo de actitudes (Nunnally, 1970). Se espera que sólo los sujetos que se sitúen en una estrecha zona alrededor de este punto estén de acuerdo con el enunciado y que los sujetos que tienen actitudes más negativas o más positivas estén en desacuerdo con él.

El procedimiento de Thurstone recurre a jueces que establecen el valor escalar de cada ítem. Cuando se utiliza la escala en estudios posteriores para medir las actitudes de otros sujetos, se le pide a cada individuo que señale los enunciados con los que está de acuerdo. La puntuación del sujeto es entonces el valor escalar medio o mediano de los ítemes aceptados.

Thurstone ideó tres métodos para medir actitudes. En todos ellos se emplean jueces para determinar el grado de relación de cada ítem con la actitud que se quiere medir. Posteriormente, los juicios emitidos se convierten en valores escalares para cada reactivo y finalmente se aplican los reactivos a los sujetos.

Uno de los problemas básicos de los métodos de Thurstone es la falta de representatividad de los jueces (Triandis, 1971). Esta es una de las razones principales por las que se emplean poco actualmente los métodos de Thurstone para la medida de la actitud, en favor de otros métodos más sencillos como el de Likert.

La *técnica del “escalograma” de Guttman* para la medida de actitudes fue desarrollada a partir de los trabajos realizados acerca de la moral de los soldados americanos después de la II Guerra Mundial (Guttman, 1944).

El modelo metodológico que sirve de base a esta escala es un *modelo monótono determinista* en el que se presupone que cada ítem tiene una relación perfecta de algún tipo con un rasgo hipotético o dimensión determinada.

Guttman propone la construcción de escalas de actitudes ordenadas. En este tipo de escalas, cuando un sujeto está de acuerdo con un enunciado (ítem), también debe estarlo con todos los ítemes de menor rango.

Un ejemplo de este tipo de escala puede ser el siguiente:	<u>SÍ</u>	<u>NO</u>
1. Las Naciones Unidas son la institución salvadora de la humanidad
2. Las Naciones Unidas constituyen nuestra mejor esperanza de paz
3. Las Naciones Unidas son una fuerza constructiva del mundo
4. Deberíamos continuar formando parte de las Naciones Unidas

En este tipo de escala existe una alta probabilidad de que cualquier persona que contestase afirmativamente al enunciado número 1, también contestara afirmativamente a los demás. Y quien no lo hiciera al 1 sino al 2, también lo hiciera al 3 y 4.

Además, con este método se logra un escalamiento conjunto de objetos y sujetos.

Las principales asunciones básicas del análisis de escalograma pueden resumirse en las siguientes (Morales, 1981; Arce, 1995):

1. Un constructo (opinión, actitud u otra característica), puede medirse a través del universo de atributos que lo caracterizan. Así, un atributo pertenece a una opinión o actitud cuando su contenido hace referencia y expresa aquello que se quiere medir.
2. Asumiendo que una actitud pueda medirse a través de enunciados que representen contenidos de opinión, pueden escalonarse las respuestas de los sujetos.
3. La técnica postula la posibilidad de establecer unidimensionalidad. Se asume que el procedimiento establece una dimensión a medir, ya que se busca la eliminación de los factores que son distintos a la dimensión que se pretende medir.
4. La principal dificultad con la que se encuentra este método es la de encontrar un conjunto de enunciados que satisfagan el modelo. Por lo que no constituye un enfoque práctico para evaluar las actitudes.

Esta técnica ha tenido una enorme difusión, hasta el punto de que muchos investigadores sociales la han considerado como el método básico para el escalonamiento de opiniones y actitudes como datos cualitativos (Morales, 1981). No obstante, en la actualidad ha sido sustituida por la técnica de Likert.

Otros métodos de escalamiento y medida de actitudes no parten del supuesto teórico común que suscriben todos los métodos anteriores, la unidimensionalidad del rasgo o constructo medido.

Los nuevos métodos hacen uso de técnicas estadísticas más complejas de tipo multivariado que posibilitan el análisis conjunto de sujetos u objetos y dimensiones de evaluación.

Las técnicas empleadas más frecuentemente en este campo son el escalamiento multidimensional (MDS) y el análisis factorial de correspondencias. Arce (1995) presenta una de las introducciones más claras a estos métodos aplicados a la medida de las actitudes.

3. TÉCNICA LIKERT PARA LA CONSTRUCCIÓN DE ESCALAS DE ACTITUD

Likert propone una técnica de elaboración de escalas que presenta un proceso de construcción más breve y sencillo que la técnica escalar de Thurstone y que es tan fiable como ésta. Se trata de la técnica más empleada para la construcción de este tipo de escalas.

El modelo metodológico que sirve de base a este procedimiento es un modelo monótono aditivo, en el que los mismos sujetos se gradúan o se escalan ellos mismos respecto al objeto de actitud. *Con el método de Likert se escalan los*

sujetos no los objetos. Son los individuos los que se sitúan en las dimensiones de actitud y no los reactivos o ítems.

El modelo aditivo supone además que los ítems individuales:

- a) tienen una relación monótona con los rasgos subyacentes (el ítem distingue bien entre los sujetos situados en un punto medio de los que poseen mucho o poco del rasgo medido), y
- b) que la suma de las puntuaciones de los ítems se relaciona de forma aproximadamente lineal con el rasgo.

Desde el punto de vista de la medición, la técnica Likert asume un nivel de medida ordinal en la que los sujetos son ordenados según su posición favorable o no hacia la actitud en cuestión.

La puntuación total de un sujeto se obtiene sumando las puntuaciones de los ítems individuales, una vez que se han invertido las puntuaciones de los enunciados de sentido negativo.

El procedimiento de Likert se aplica tanto a los ítems dicotómicos como a los ítems de muchos puntos, aunque se emplea con mucha mayor frecuencia en estos últimos.

3.1. FASES EN LA CONSTRUCCIÓN DE LA ESCALA DE LIKERT

La construcción de una escala tipo Likert supone las siguientes fases (Morales, 1981; Arce, 1995):

1. Preparación de los ítems iniciales
2. Administración de ítems a una muestra de sujetos
3. Asignación de los puntajes a los ítems
4. Asignación de las puntuaciones a los sujetos
5. Análisis y selección de los ítems
6. Análisis de la fiabilidad de la escala
7. Preparación de la aplicación de la versión final de la escala

3.1.1. Preparación de los ítems iniciales

Inicialmente se formulan una serie de enumerados o proposiciones que ha de ser superior a los que aparezcan en la escala final. Lo usual es que, si la escala final contiene 5 a 8 ítems por cada dimensión homogénea, se parta de un número mayor de enunciados.

Los enunciados han de referirse al tema o actitud que se pretende medir, y es conveniente tener un conocimiento profundo sobre el tema objeto de estudio. Deben cubrirse los aspectos clave del objeto que se va a evaluar.

En la selección de las proposiciones Likert aconseja, además:

1. Que los enunciados sean expresión del comportamiento deseado y no de hecho. Se trata de formular los enunciados en forma categórica y en sen-

tido del deber ser. Así, por ejemplo, los enunciados deben tener esta forma: “Me siento mal cada vez que tengo un examen”, “Estoy satisfecho con mi vida”, o “Un matrimonio mal avenido debería disolverse por mutuo acuerdo”.

2. Los enunciados deben ser claros, concisos y directos.
3. Para evitar las respuestas estereotipadas, los ítems deben combinarse de tal modo que el 50% de ellos correspondan a la zona positiva de la actitud y la otra mitad a la negativa.
4. Los enunciados deben implicar una sola variable de actitud. Este supuesto es muy importante ya que si se parte del hecho de que estamos midiendo un objeto unidimensional, se necesitan escalas homogéneas.

3.1.2. Administración de ítems

Los ítems elaborados inicialmente se aplican a una muestra suficientemente representativa de los sujetos cuyas actitudes deseamos medir. La diferencia existente entre este paso y el de Thurstone, está en que los enunciados no se someten a una evaluación de jueces expertos, sino que se ponen a prueba en parte del colectivo que se pretende estudiar. En la técnica de Likert los jueces tienen un peso más importante en la formulación inicial de los enunciados.

Los ítems se presentan en un cuestionario o escala de modo que para cada uno de ellos hay distintas alternativas graduadas en intensidad:

Absolutamente de acuerdo	—
De acuerdo	—
Indiferente	—
En desacuerdo	—
Absolutamente en desacuerdo	—

Hay que prestar atención al hecho de que la forma de la respuesta depende del contenido del ítem. Así, en algunos casos hay que referirse a la frecuencia:

Mucho	—
Bastante	—
Regular	—
Poco	—
Nada	—

La escala numérica debe definirse explícitamente, como por ejemplo:

- 1= Totalmente en desacuerdo
- 2= En desacuerdo
- 3= Ni acuerdo ni desacuerdo
- 4= De acuerdo
- 5= Totalmente de acuerdo

La presentación material de la escala suele combinar una escala gráfica con una numérica. Es preferible emplear una escala gráfica con números que emplear números sin escala gráfica. La escala gráfica debería disminuir los errores que se producen al registrar las evaluaciones. Si los significados de los números se dan únicamente al comienzo del inventario y los sujetos deben recordarlos al registrar los números en los espacios en blanco, es probable que los olviden o confundan.



Otra forma de presentar la escala es con intervalos cerrados, como puede ser:



La adscripción de valores a las respuestas se deja a elección del investigador. Así, la numeración puede comenzar en cero o en uno, se puede dejar el cero para el punto de indiferencia, etc. Por otra parte, en las escalas aditivas no es necesario que haya un punto neutro.

Otro tipo de punto de referencia para las escalas de evaluación se refiere a la conducta real de las personas. La siguiente es una escala numérica relativa a la tendencia de los pacientes mentales a tener alucinaciones:

- 0= No muestra signo alguno de tener alucinaciones.
- 1= Ocasionalmente gesticula y habla solo como si tuviera alucinaciones, pero no las verbaliza.
- 2= Verbaliza alucinaciones sólo ocasionalmente.
- 3= Verbaliza activamente alucinaciones la mayor parte del tiempo.

El empleo de puntos de referencia relacionados con la conducta presenta, sin embargo, algunas dificultades. En primer lugar, porque se necesita un conjunto diferente de puntos de referencia para cada escala lo que hace difícil la construcción de inventarios basados en estas escalas.

3.1.3. Asignación de los puntajes a los ítems

En esta fase hay que clasificar a cada ítem según su posición positiva o negativa de actitud. Además hay que ponderar las distintas alternativas de respuesta a los ítems. Sobre las formas de puntuar las categorías de respuesta existen diversos criterios. Nos encontramos con graduaciones de 3, 4, 5, 6 ó 7 categorías de respuesta. Aunque el número de graduaciones de respuesta puede depender de la

naturaleza y objetivos de la investigación, muchos investigadores se inclinan por escalas con 5 o 7 puntos.

La decisión sobre el número de grados que debe tener una escala es un aspecto importante. En términos de la teoría psicométrica es preferible utilizar un número mayor de grados. Muchos estudios prueban que la fiabilidad de las escalas individuales de evaluación está en función directa del número de grados de la escala (Nunnally, 1970).

A medida que el número de grados de la escala pasa de 2 a 20 la fiabilidad aumenta, en un principio de forma muy rápida. El aumento tiende a estabilizarse hacia los siete grados y más allá de los 11 grados la fiabilidad mejora muy poco. Esto parece contrario al sentido común, pues se podría pensar que con más grados el sujeto tiene más dónde decidir y hay más posibilidades de que en el retest señale un número distinto. Aunque esto es así y con más grados la varianza de error aumenta, la varianza del puntaje verdadero aumenta mucho más.

En cuanto a la decisión a adoptar sobre el número par o impar de grados, parece que no hay razones claras para utilizar uno u otro número; sin embargo, algunos autores indican la ventaja de tener un número par de grados para evitar sesgos de respuesta intermedios.

Aunque el número de grados a utilizar es importante si se emplea una sola escala, cuando se suman las puntuaciones de varias escalas, 6 al menos, como sucede en las escalas tipo Likert, su importancia es menor, ya que la fiabilidad de la escala está muy en relación con el número de elementos que componen la escala y la correlación que hay entre ellos.

3.1.4. Asignación de las puntuaciones a los sujetos

La puntuación total de un individuo en la escala es la suma de sus puntuaciones a cada ítem. La intensidad de actitud y su valoración depende del número de ítems y de los puntajes asignados, en una medida que está en relación con la amplitud de la variable que se quiere medir. Así, si tenemos 20 ítems que han sido ponderados cada uno con valores que van de 0 a 6, la amplitud de la variable es de un mínimo de 0 puntos y un máximo de 120.

3.1.5. Análisis y selección de los ítems

Establecidas las puntuaciones totales de la muestra de sujetos, corresponde establecer un análisis de las respuestas para determinar qué ítems son más adecuados para formar parte de la escala definitiva.

Existen varios sistemas para la selección de los ítems:

1) Cálculo del poder discriminativo de cada ítem.

Parte del cálculo del valor promedio de cada ítem para los grupos con puntuaciones altas y bajas en el total de la escala (normalmente el 27% superior y el

27% inferior). Con los valores promedios se procede al cálculo del índice de discriminación, mediante la prueba estadística t de Student. Los ítems seleccionados son los que evidencian diferencias significativas entre ambos grupos.

2) Correlación ítem-test.

Este método requiere el cálculo del coeficiente de correlación de cada ítem con el total de la escala. En este caso, el supuesto de que se parte es que la distribución es normal y continua. Se usa, generalmente, el coeficiente de correlación biserial (r_b) puesto que supone la covariación entre una variable de intervalo y otra dicotómica establecida por el investigador.

El grado de integración de un ítem se puede descubrir investigando la relación que existe entre las respuestas que los sujetos emiten para dicho ítem y las respuestas que los sujetos emiten para el conjunto total de ítems. Lógicamente, una relación baja (o negativa) será una indicativa de que el ítem no está bien integrado. Por contra, una relación moderada o alta indica que la integración del ítem es aceptable o excelente.

3) Test de la mediana.

Es la mejor alternativa de las presentadas hasta ahora, ya que se corresponde más claramente con el nivel ordinal de medición.

Hay que determinar, en primer lugar, el valor de la mediana de cada ítem para los grupos bajo y alto en forma combinada. Con ello, preparamos una tabla de contingencia 2x2 para cada ítem a la que luego aplicaremos pruebas de significación adecuadas, como la prueba χ^2 . La tabla de contingencia quedaría representada de la siguiente manera:

	Grupo alto	Grupo bajo
Puntajes por debajo de la mediana	A	B
Puntajes por encima de la mediana	C	D

Tabla 1. *Tabla de contingencia entre las puntuaciones de los grupos alto y bajo en el total de la escala y las respuestas a un ítem.*

Pueden utilizarse dos pruebas de significación:

a) Ji-cuadrado (χ^2)

$$\chi^2 = \frac{N[(AD-BC)-N/2]^2}{(A+B)(C+D)(A+C)(B+D)} ;$$

Su empleo es adecuado cuando N es igual o mayor a 40.

b) Test de probabilidad exacta de Fisher

Cuando el número de sujetos en una celdilla es menor del 5% del total de sujetos con los que se calcula la prueba.

$$P = \frac{(A+B)! (C+D)! (A+C)! (B+D)!}{N! A! B! C! D!};$$

Con estas pruebas estadísticas se seleccionan aquellos ítems que presentan diferencias significativas, rechazando aquellos otros que sean poco discriminativos.

4) Empleo del análisis factorial

Este procedimiento sirve para establecer la contribución de cada ítem a la escala total, a través de la observación de las saturaciones o cargas factoriales de cada uno de ellos. Es adecuado también para establecer la homogeneidad de la escala o la existencia de factores homogéneos dentro de la escala total. Aunque se trata de uno de los métodos más empleados en la actualidad para el análisis y la selección de ítems (enunciados) hay autores que consideran más apropiados otros métodos como el análisis de fiabilidad (Nunnally, 1970)

5) Contribución de cada ítem a la fiabilidad total de la escala

Algunas técnicas de estimación de la fiabilidad de las escalas como el coeficiente de consistencia interna alpha de Cronbach, además de establecer la contribución que hace cada ítem a la fiabilidad de la escala total, indican el valor de la fiabilidad de la escala cuando se elimina cada uno de los elementos de la misma. Esto ofrece la posibilidad de decidir sobre el mantenimiento de aquellos enunciados que contribuyen a mantener un determinado valor del coeficiente de fiabilidad y de aquellos otros enunciados cuya eliminación supone una mejora de la fiabilidad total de la escala.

3.1.6. Establecimiento de la fiabilidad de la escala

El análisis de ítems ayuda a obtener una escala compacta, en la que todos los ítems estén integrados y contribuyan a la medida del constructo subyacente. Además, el análisis de ítems ayuda a identificar ítems que contienen información redundante. Si descubrimos que dos o más ítems contienen casi la misma información, bastaría con quedarse con uno de ellos en la escala, dado que los demás contienen información repetida y son, en cierto modo, inútiles.

En la tabla 2 se ofrece la correlación “corregida” entre cada ítem de una escala de un total de cinco ítems y la puntuación en el total de la escala.

ITEM	Correlación
Ítem 1	0,78
Ítem 2	0,69
Ítem 3	0,65
Ítem 4	0,75
Ítem 5	0,69

Tabla 2. *Correlación entre cada ítem y el total de la escala*

Se observa que todos los ítems tienen una correlación moderada, aceptable, con el total, mientras que en la tabla siguiente, tabla 3, se ofrece la matriz de correlaciones entre los 5 ítems. Una vez que las correlaciones son moderadas, se puede afirmar que los ítems miden distintos matices del mismo constructo.

	1	2	3	4	5
1	1.0000				
2	.3857	1.0000			
3	.5363	.3823	1.0000		
4	.6231	.4435	.5460	1.0000	
5	.4874	.3311	.4654	.3742	1.0000

Tabla 3. *Matriz de correlaciones hipotéticas para los 5 ítems.*

La fiabilidad puede calcularse mediante distintos procedimientos, como el test –retest, la división por la mitad o el cálculo de la consistencia interna. Para escalas homogéneas, una forma muy aceptada consiste en calcular el coeficiente alpha de consistencia interna (α) de Cronbach (1951), que puede definirse de la siguiente manera:

$$\alpha = \frac{nr}{1 + r(n-1)}$$

donde:

n es el número de ítems;

r es la media de las correlaciones entre los ítems.

Con los datos del ejemplo, la media de las correlaciones entre los 5 ítems es .4575. Por tanto,

$$\alpha = \frac{5 \times 0.4575}{1 + (0.4575 \times 4)} = .81$$

En este caso el coeficiente alpha es bastante aceptable (valor máximo 1), por lo que puede considerarse la fiabilidad de la escala como buena. El valor α , además de depender de la correlación entre los ítems, depende del número de ítems. Para una correlación media dada, aumentando el número de ítems, a 10 por ejemplo, alpha alcanzaría un valor mayor (.89). Sin embargo, si el valor actual de la escala se considera suficientemente alto (.81), no tiene mucha justificación aumentar el tamaño de la escala, aunque así se consiga un aumento en su fiabilidad.

El análisis de fiabilidad también puede utilizarse para mejorar la fiabilidad de la escala, eliminando aquellos ítems que no contribuyen o lo hacen de forma que disminuyen la fiabilidad de la misma. Esto se consigue mediante una fórmula que establece el aumento de fiabilidad esperada cuando se elimina un elemento que muestra una relación débil o negativa con los demás. Este tipo de resultado aparece en programas como el SPSS.

3.1.7. Preparación de la aplicación de la versión final de la escala

Después de rechazar los ítems inadecuados, se procede a la construcción de la versión final de la escala. Se dispondrán los ítems en un formulario con su numeración correspondiente y con las categorías de respuesta. La escala ha de contener las instrucciones necesarias para su aplicación, la explicación de lo que se quiere administrar, la demanda de colaboración y las indicaciones para la respuesta.

En la versión final hay que considerar que las puntuaciones definitivas que se adscriben a los sujetos serán la suma de las puntuaciones obtenidas en cada ítem; o bien esta suma dividida entre el total de los ítems.

3.2. VENTAJAS Y DESVENTAJAS DE LA TÉCNICA LIKERT

Al igual que la técnica de Thurstone, el método de Likert presenta un conjunto de ventajas y desventajas que suelen presentarse frecuentemente, en referencia con las del primer autor.

El investigador debe conocer no sólo los fundamentos o criterios en que se basa la construcción, sino también los inconvenientes principales que presenta, pues le servirá tanto para orientar sus propios criterios de medición como para establecer su propia economía de medios y esfuerzos.

Entre las principales ventajas y desventajas que se han señalado (Morales, 1981) figuran:

a) Ventajas:

1. Permite la utilización de ítems que no están directamente relacionados con la actitud que se pretende medir. En tal sentido, se da la posibilidad de establecer mediciones de carácter indirecto que en el método Thurstone se halla seriamente limitado.
2. Su construcción es sencilla. Puede realizarse con un trabajo menor y de modo más rápido que en las escalas Thurstone.
3. El número de ítems que se necesitan para su confección suele ser menor que en las de Thurstone. Su fiabilidad suele ser muy elevada, permitiendo, además, el establecimiento de graduaciones o intensidad de respuesta.
4. Las posibles respuestas a cada ítem permiten un mayor grado de información sobre la actitud estudiada.

b) Desventajas

1. En la escala tipo Likert, la magnitud comparativa no expresa en qué medida es más favorable la actitud de un sujeto respecto de otro. Esto se debe a que el nivel de medida es ordinal.
2. La puntuación de un sujeto suele tener un significado algo confuso, ya que se puede obtener el mismo puntaje de actitud con ítems que tienen contenido diverso. Esta crítica se ha formulado igualmente a la técnica Thurstone. Parece que este problema no es serio cuando el sentido de la actitud es congruente, pero da lugar a dificultades cuando se trata de determinar si la puntuación de un individuo puede tener significados diferentes.
3. Una cuestión muy importante es la garantía de unidimensionalidad. No obstante, las técnicas de análisis de ítems como el coeficiente de consistencia interna α y el análisis factorial, sirven para contrastar empíricamente este supuesto.

4. EL ANÁLISIS FACTORIAL EN LA CONSTRUCCIÓN DE LAS ESCALAS DE ACTITUD

Entre las técnicas correlacionales de tipo multivariado, el análisis factorial es uno de los procedimientos más utilizados. La finalidad principal de este tipo de análisis es reducir la información existente sobre la covariación de un conjunto de variables, agrupando aquellas variables que tengan una varianza común en uno o más factores subyacentes. O de otra manera, dado un conjunto de variables linealmente relacionadas (n), el objetivo principal del análisis factorial es centrarse en lograr la reducción de éstas a un número menor de factores (k) que per-

mitan describir las intercorrelaciones entre las variables. Estos factores pueden estar o no correlacionados entre sí, aunque generalmente la propia lógica del análisis factorial establece que sean lo más independientes posible.

La necesidad del empleo del análisis factorial se deriva de esclarecer y resumir las relaciones subyacentes a un gran número de variables. Pensemos por ejemplo en los ítems de un cuestionario o en las posibles medidas del rendimiento académico. Un cuestionario con 30 ítems únicamente da lugar a 415 intercorrelaciones distintas. En estas intercorrelaciones puede haber un único factor subyacente a todas ellas, indicándonos que todas miden fundamentalmente lo mismo o pueden existir 415 factores si no aparece ningún factor común entre las variables.

Cualesquiera que sean los objetivos de un análisis factorial, implica una serie de pasos desde la medida de las variables hasta la extracción de factores. Estos pasos son los siguientes (Comrey, 1985):

- a) seleccionar las variables;
- b) calcular la matriz de correlaciones entre las variables;
- c) extraer los factores siguiendo un método que esté de acuerdo con nuestros objetivos;
- d) rotar los factores de manera que sean más interpretables los resultados;
- e) interpretar la matriz de los factores rotados.

No es nuestra intención en este tema presentar los distintos procedimientos de extracción y rotación de factores debido entre otras cosas a la dificultad que conlleva este cálculo, que normalmente requiere el empleo de programas estadísticos informatizados. Se trata únicamente de presentar las bases conceptuales para la comprensión general de este método de análisis de datos.

El análisis factorial comienza con una matriz de coeficientes de correlación entre los datos de las variables bajo estudio. Un ejemplo lo tenemos cuando aplicamos distintos test de personalidad a un grupo más o menos amplio de estudiantes. El número de sujetos está en función entre otras cosas del número de variables utilizadas, considerándose que en ningún caso debe ser inferior al número de variables multiplicado por cinco, siempre y cuando este valor supere como mínimo los 50 sujetos.

En la tabla 4 aparece la matriz simétrica de correlaciones (hipotética) entre 6 ítems de un cuestionario sobre personalidad.

Después que se ha calculado la matriz de correlaciones R , el siguiente paso es la extracción de factores. Este proceso implica un procedimiento numérico que usa la totalidad de los coeficientes de la matriz R para producir un vector de coeficientes que relacionan las variables medidas con una construcción hipotética denominada factor. En términos matemáticos, se trata de encontrar una matriz de factores F que multiplicada por su traspuesta reproduzca las correlaciones ori-

ginales entre los datos y que contenga un menor número de factores que de variables. El procedimiento general consiste en extraer factores de la matriz de correlaciones R hasta que no exista varianza residual apreciable.

VARIABLES (ÍTEMS)	1	2	3	4	5	6
1	1					
2	,72	1				
3	,01	,16	1			
4	,27	,40	,64	1		
5	,00	,12	,48	,48	1	
6	,63	,56	,00	,21	,00	1

Tabla 4. Matriz de correlaciones entre los seis ítems de un cuestionario

Hay muchos métodos de extracción de factores, pero todos ellos finalizan con una columna de números, uno para cada variable, que indica los “pesos”, “cargas” o “saturaciones” de las variables en ese factor. Estos pesos representan el grado en que cada variable se relaciona o está representada en el factor hipotético. En la mayoría de los métodos de extracción de factores, estos pesos indican la correlación entre las variables y el factor.

El número de factores extraídos dependen del porcentaje de varianza explicada por cada factor. Esta varianza explicada viene establecida por la raíz latente asociada al factor correspondiente, que está definida por la suma del cuadrado de los pesos factoriales de ese factor. En la práctica, se suele tomar como criterio para mantener un factor en el análisis el que su raíz latente sea igual o mayor que uno.

El modelo de extracción de factores más empleado, posiblemente, sea el modelo de factor principal, que tiene a su vez dos variaciones: el método de componentes principales y el método centroide.

Con el *método de componentes principales* de Hotelling se obtienen los factores de manera que el primer factor recoge la máxima varianza posible, y así sucesivamente. El primer factor es la combinación lineal de las variables que hacen máxima la cantidad de varianza explicada de la matriz R; el segundo, contribuye a minimizar la varianza residual y a explicar el máximo de la varianza restante, y así sucesivamente. Además, los factores extraídos no correlacionan entre sí.

El *método centroide* formulado por Thurstone hace mínimas las distancias de las variables a un centro de gravedad representado por la media de los valores

que representan a las variables que forman parte de la matriz de correlación. El método centroide, por tanto distribuye la varianza total explicada entre los factores de forma que aunque los primeros factores que surgen en el análisis explican un mayor porcentaje de varianza, estos factores no hacen máxima la varianza explicada por cada uno de ellos. Se sigue una regla de distribución de la varianza entre los factores, más que de acumulación de esta varianza en los primeros factores que aparecen en el análisis. El procedimiento de cálculo para la extracción de factores es más sencillo en el método centroide que en el de componentes principales.

Una vez que se extraen los factores necesarios para explicar las correlaciones de la matriz, estos pesos se ordenan en una tabla denominada matriz factorial no-rotada. Para los datos del ejemplo anterior, los factores encontrados son los que aparecen en la tabla 5.

El primer factor de la tabla 5 es el mayor, una vez que la raíz latente o suma de los cuadrados de los pesos factoriales es la más alta. Esto es, el primer factor es el que más varianza explica (2,27 respecto a 3,71, es igual a 61,3%). El factor II explica el 38,7 de la varianza total explicada.

La última columna de la matriz de la tabla 3 contiene las comunalidades h^2 para las variables. Las comunalidades son iguales a las sumas de los cuadrados de las cargas factoriales de las variables en los dos factores. Así, la comunalidad para la variable 1 es igual a ,81 ($,76^2 + ,60^2$). Las comunalidades indican la cantidad de solapamiento entre las variables y los factores. O lo que es lo mismo, la cantidad de la varianza de la variable que se explica por todos los factores.

El análisis factorial continúa con la rotación de los factores. Este procedimiento tiene la finalidad de facilitar la interpretación de los factores obtenidos una vez que mediante la rotación se logra cada variable saturar sobre el menor número de factores posible y que haya el mayor número de ceros posible en la matriz rotada. Esto facilita la obtención de factores “puros” en el sentido que las variables que saturan en un factor tienden a no hacerlo en los otros.

Hay dos métodos principales de rotación, la rotación ortogonal, que hace uso del método varimax, y la rotación oblicua. La rotación ortogonal va destinada a obtener factores no correlacionados, mientras que la rotación oblicua permite la correlación entre los factores. El método de rotación varimax, ideado por Kaiser, maximiza la varianza entre los factores. Es el método ortogonal más utilizado. Otros métodos de rotación ortogonal, como el método quartimax, tienden a encontrar un factor general.

VARIABLES (ÍTEMS)	Factor I	Factor II	Comunalidad h^2
1	,67	-,60	81
2	,73	-,38	,68
3	,53	,59	,64
4	,75	,40	,73
5	,40	,45	,36
6	,52	-,46	,49
	λ 2,27	1,44	3,71

Tabla 5. Matriz factorial no-rotada.

En la tabla 6 aparece la matriz factorial rotada, siguiendo el método varimax, correspondiente a los datos del ejemplo que venimos comentando. Como podemos ver, aparecen de forma más nítida los dos factores. El primero, con un 61,3% de la varianza explicada, está definido por las variables (ítems) número 1, 2 y 6; mientras que el segundo está definido por los ítems 3, 4 y 5. En función del contenido de estos ítems podemos denominar el factor.

VARIABLES (ÍTEMS)	Factor I	Factor II	Comunalidad h^2
1	,90	,00	81
2	,80	,20	,68
3	-,00	,80	,64
4	,30	,80	,73
5	-,00	,60	,36
6	,70	,00	,49
	λ 2,03	1,68	3,71

Tabla 6. Matriz factorial rotada, método varimax.

Estas transformaciones, sin embargo, respetan la solución factorial, por lo que las variables conservan el mismo valor de sus comunalidades y los factores aportan la misma explicación de la varianza, en conjunto.

Para finalizar este apartado nos referiremos a la prueba de esfericidad de Bartlett, cuya finalidad es la de establecer si existen correlaciones distintas de 0 en

la matriz de correlaciones inicial, supuesto necesario para la realización de un análisis factorial, y que debe de utilizarse antes de comenzar el proceso de análisis.

5. FASES EN LA ELABORACIÓN DE UNA ESCALA. EJEMPLO DE CONSTRUCCIÓN DE UNA ESCALA DE EVALUACIÓN DE LA DOCENCIA UNIVERSITARIA (Universidad de La Laguna)

5.1. ESTABLECIMIENTO DE LAS DIMENSIONES MÁS COMUNES DE LAS ESCALAS DE EVALUACIÓN DEL DESEMPEÑO DOCENTE (REVISIÓN DE LA LITERATURA CIENTÍFICA)

1. Cumplimiento de obligaciones docentes: Puntualidad y asistencia a clases y tutorías.
2. Organización del curso y estructura y claridad de las clases
3. Conocimiento y dominio de la materia.
4. Motivación y entusiasmo del profesor
5. Relación profesor-alumno: Comunicación, participación, motivación e interés
6. Uso de recursos materiales y didácticos
7. Clases prácticas: Suficiencia y relación con la teoría
8. Evaluación: Normas, criterios y procedimientos
9. Sobrecarga de trabajo percibida por el alumnado
10. Valoración global.

5.2. ELABORACIÓN DEL CUESTIONARIO

Cuestionario inicial que contiene 34 afirmaciones referidas a las 10 dimensiones teóricas identificadas.

Versión inicial sometida al criterio de jueces expertos en la materia sobre la que versa el cuestionario.

Formato de respuesta de 4 alternativas: 1= Nada característico, 2= Algo característico, 3= Bastante característico, y 4= Muy o totalmente característico.

Número de componentes de la muestra N= 5350 alumnos/as universitarios.

5.3. RESULTADOS DEL ANÁLISIS FACTORIAL (MÉTODO DE COMPONENTES PRINCIPALES Y ROTACIÓN ORTOGONAL-VARIMAX)

FACTOR I: DOMINIO (36% varianza explicada)

ITEMS
7. El profesor/a ilustra adecuadamente los conceptos teóricos (.66)
9. La forma en que el profesor/a da clase parece adecuada a las características de estegrupo y esta materia (.67)
10. El profesor/a domina la materia (.74)
11. Con frecuencia cita referencias, demostraciones, técnicas y/o avances recientes en la materia (.56)
12. En general, creo que el profesor/a responde con rigor y precisión a lo que se le consulta (.56)
32. Hemos aprendido mucho asistiendo a sus clases (.66)
34. Teniendo en cuenta las limitaciones del medio en que se desenvuelve, creo razonadamente que es un buen profesor/a (.65)

() Entre paréntesis se ofrece el valor de las saturaciones factoriales.

FACTOR II: ORGANIZACIÓN/PLANIFICACIÓN y EVALUACIÓN (6% varianza explicada)

ITEMS
5. Al comienzo del curso, el profesor/a da toda la información necesaria sobre el programa y el plan de trabajo (.55)
20. Durante las horas de tutoría, atiende correctamente las consultas de los estudiantes (.59)
25. Las evaluaciones se ajustan a lo trabajado durante el curso (.54)
26. Las normas y actividades de evaluación las da a conocer desde el principio del curso (.69)
27. Sus criterios de evaluación son claros y pertinentes (.64)
28. Informa en un tiempo razonable de los resultados de la evaluación (.60)
29. Comenta los resultados de las evaluaciones, incluso individualmente, con los alumnos que lo solicitan (.65)

() Entre paréntesis se ofrece el valor de las saturaciones factoriales.

FACTOR III: MOTIVACIÓN (4.8% varianza explicada)

ITEMS
13. Se preocupa por fomentar el espíritu crítico entre los alumnos/as (,73)
14. El profesor/a trata de despertar en nosotros el interés por la materia (,60)
16. Normalmente, mantiene una actitud receptiva y respetuosa ante las preguntas y sugerencias de los alumnos (,42)
17. El profesor no fomenta la participación en clase (,50)
18. Facilita la libre expresión de ideas entre los estudiantes (,72)
19. Anima a la discusión académico-científica de los contenidos abordados (,77)

() Entre paréntesis se ofrece el valor de las saturaciones factoriales.

FACTOR IV: UTILIZACIÓN DE MEDIOS Y RECURSOS DIDÁCTICOS (4.5% varianza explicada)

ITEMS
21. Los recursos disponibles (libros, artículos, material e instrumental técnico) le facilitan la impartición de la materia (,45)
22. Utiliza con regularidad diversos materiales y recursos didácticos en clase (pizarra, video, retroproyector, diapositivas, láminas, ordenador o recursos propios de la materia (,70)
23. Las clases prácticas de la materia (laboratorios, talleres, estudios clínicos, seminarios, etc...) complementan eficazmente a la teoría (,81)
24. Las prácticas (demostraciones, ensayos, etc) se ilustran con contenidos reales (,75)

() Entre paréntesis se ofrece el valor de las saturaciones factoriales.

FACTOR V: CUMPLIMIENTO DE OBLIGACIONES DOCENTES (3.6% varianza explicada)

ITEMS
El profesor/a asiste a clase regularmente, sin faltas injustificadas (,81)
Cuando por motivos personales el profesor/a pierde clases, las recupera (,65)
El profesor/a suele ser puntual en las horas de clase (,75)

() Entre paréntesis se ofrece el valor de las saturaciones factoriales.

FACTOR VI: VALORACIÓN NEGATIVA DEL DESEMPEÑO DOCENTE (3.4% varianza explicada)

ITEMS
4. Con frecuencia, el profesor/a es muy irregular en el cumplimiento del horario de tutoría (,66)
8. No organiza bien las clases (teóricas y/o prácticas) (,54)
15. En general, el profesor/a no disfruta dando clases (,52)
33. En general, no estoy satisfecho/a del trabajo que ha llevado/está llevando a cabo este curso (,44)

() Entre paréntesis se ofrece el valor de las saturaciones factoriales.

FACTOR VII: EXIGENCIA (3% varianza explicada)

ITEMS
30. Con frecuencia, los alumnos/as nos quejamos de que nos exige demasiado (,82)
31. Necesitamos invertir demasiado tiempo si queremos aprobar la materia que imparte este profesor/a (,83)

() Entre paréntesis se ofrece el valor de las saturaciones factoriales.

Los resultados obtenidos con el cuestionario indicaron que *la percepción de una mayor exigencia y dificultad en la materia por parte del estudiante no conllevó una valoración negativa (ni positiva) del desempeño docente del profesorado.*

Una mayor exigencia y dificultad se percibe junto a una mayor asistencia y puntualidad (cumplimiento del profesor).

5.4. FIABILIDAD DE CONSISTENCIA INTERNA DE LOS FACTORES (N= 5350)

FACTORES	ALPHA (α de Cronbach)
DOMINIO	0,86
ORGANIZACIÓN y EVALUACIÓN	0,85
MOTIVACIÓN	0,87
UTILIZACIÓN DE RECURSOS	0,74
PUNTUALIDAD Y CUMPLIMIENTO	0,69
VALORACION NEGATIVA DEL DOCENTE	0,67
EXIGENCIA	0,74

5.5. VERSIÓN FINAL DE LA ESCALA

ESCALA DE EVALUACIÓN DE LA CALIDAD DE LA DOCENCIA

Cuadernillo para Estudiantes

Este cuestionario recoge información acerca de cómo desarrollan su actividad docente los profesores de la Universidad de La Laguna. Con los datos obtenidos se realizará y publicará un informe sobre la docencia en la Universidad. Dada la importancia de estos objetivos, es de suma trascendencia que respondas con la máxima responsabilidad y sinceridad.

Hay que completar un cuestionario para cada profesor o profesora.

El cuestionario está formado por 16 frases que describen el quehacer docente del profesor, a las que se contesta en una escala de 0 a 9 puntos.

Marca el 0 si crees que la frase no se aplica en absoluto al comportamiento de este profesor.

Marca el 9 si crees que la frase se aplica totalmente a este profesor. Matiza tu respuesta utilizando los puntos intermedios de la escala. Procura contestar a todas las preguntas.

GRACIAS POR TU COLABORACIÓN

NOMBRE DEL PROFESOR
 ASIGNATURA.....
 LICENCIATURA/DIPLOMATURA
 CENTRO
 CURSO GRUPO.....

CUESTIONARIO DE DESEMPEÑO DOCENTE

0. ¿Has asistido a más del 50% de las clases del Profesor?

SÍ

NO

	NO APLICABLE EN ABSOLUTO A ESTE PROFESOR					TOTALMENTE APLICABLE A ESTE PROFESOR				
	0	1	2	3	4	5	6	7	8	9
1. El profesor/a domina la materia.....	0	1	2	3	4	5	6	7	8	9
2. Con frecuencia, cita referencias, demostraciones, técnicas y/o avances recientes en su materia.....	0	1	2	3	4	5	6	7	8	9
3. En general, creo que el profesor/a responde con rigor y precisión a lo que se le consulta.	0	1	2	3	4	5	6	7	8	9
4. Al comienzo del curso, el profesor/a da toda la información necesaria sobre el programa y el plan de trabajo.....	0	1	2	3	4	5	6	7	8	9
5. Sus criterios de evaluación son claros y pertinentes.....	0	1	2	3	4	5	6	7	8	9
6. Comenta los resultados de las evaluaciones, incluso individualmente, con los alumnos/as que lo solicitan.....	0	1	2	3	4	5	6	7	8	9
7. Se preocupa por fomentar el espíritu crítico entre los alumnos/as.....	0	1	2	3	4	5	6	7	8	9
8. El profesor/a trata de despertar en nosotros el interés por la materia.....	0	1	2	3	4	5	6	7	8	9
9. Normalmente, mantiene una actitud receptiva y respetuosa ante nuestras preguntas y sugerencias.....	0	1	2	3	4	5	6	7	8	9
10. Utiliza con regularidad diversos materiales y recursos didácticos en clase (p.e., pizarra, video, retroproyector, diapositivas, láminas, ordenador o recursos propios de la materia)	0	1	2	3	4	5	6	7	8	9

11. Las clases prácticas de la materia (laboratorio, talleres, estudios clínicos, seminarios, etc...), complementan eficazmente a la teoría.....	0	1	2	3	4	5	6	7	8	9
12. Las prácticas (demostraciones, ensayos), se ilustran con contenidos reales	0	1	2	3	4	5	6	7	8	9
13. El profesor/a asiste a clase regularmente, sin faltas justificadas	0	1	2	3	4	5	6	7	8	9
14. Cuando por motivos personales el profesor/a pierde clases, las recupera	0	1	2	3	4	5	6	7	8	9
15. Con frecuencia, los alumnos/as nos quejamos de que nos exige demasiado	0	1	2	3	4	5	6	7	8	9
16. Necesitamos invertir demasiado tiempo si queremos aprobar la materia que imparte	0	1	2	3	4	5	6	7	8	9

BIBLIOGRAFÍA

- ARCE, R. (1995). *La medida de las actitudes*. Madrid: Síntesis.
- COMREY, A.L. (1985). *Manual de análisis factorial*. Madrid. Cátedra.
- EDWARDS, A.L. y KILPATRICK, F.P. (1948). A technique for the construction of attitude scale. *Journal of Applied Psychology*, 32, 374-384.
- GUTTMAN, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, 9, 139-150.
- LIKERT, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 1-55.
- MORALES, J.F. (1981). Las escalas de actitudes. En J.F. Morales (Ed.), *Metodología y Teoría de la Psicología*. Madrid: UNED.
- NUNNALLY, J.C. (1970). *Introducción a la medición psicológica*. Buenos Aires: Paidós.
- THURSTONE, L.L. (1931). The measurement of social attitudes. *Journal of Abnormal and Social Psychology*

