

Metodologías científicas en Psicología

Metodologías científicas en Psicología

Orfelio G. León García

Ignacio Montero García-Celay

Título original: Metodologías científicas en Psicología
Diseño del libro, de la cubierta y de la colección: Manel Andreu

Primera edición en lengua castellana: Mayo 2006

Orfelio G. León García, Ignacio Montero García-Celay

© 2006 Editorial UOC
Avda. Tibidabo, 47 08035-Barcelona
www.editorialuoc.com

Realización editorial: Editorial UOC ha generado este libro con tecnología XML / XSL.
Impresión: Gráficas Rey, SL
ISBN: 84-9788-316-0
Depósito legal: B-30307-2006

La UOC queda facultada expresamente por el/la autor/a para digitalizar y publicar la Obra en un repositorio en línea que será accesible al público bajo licencias Creative Commons, incluyendo la licencia ReconocimientoNoComercial-SinObraDerivada (BY-NC-ND), v.4.0 Internacional (jurisdicción internacional), que permitirá copiar, distribuir y transmitir públicamente la Obra siempre citando la autoría y la fuente, sin hacer un uso comercial y sin hacer obra derivada. Si la Obra es transformada, la obra generada estará sometida a una licencia Creative Commons similar o compatible con la licencia mencionada..

Autores

Orfelio G. León García

Doctor en Psicología. Profesor titular del Área de Metodología de las Ciencias del Comportamiento de la Universidad Autónoma de Madrid.

Ignacio Montero García-Celay

Doctor en Psicología. Profesor titular del Área de Metodología de las Ciencias del Comportamiento de la Universidad Autónoma de Madrid.

*Para Montse Vall-Llobera... y para todos los que, como ella, han apreciado nuestro modo de
presentar la Metodología.*

Índice

Prólogo	15
---------------	----

Capítulo I. Introducción y metodología de encuestas	17
--	-----------

1. La conexión epistemológica	17
1.1. Presentación: un relato ilustrativo	17
1.2. Diferentes formas de conocimiento	19
1.3. Características de la ciencia como actividad que genera conocimiento	22
1.4. Cómo se genera y organiza el conocimiento científico	24
1.5. La ciencia de la Psicología	27
2. El proceso de contraste de hipótesis	30
2.1. Proceso de investigación	30
2.2. El problema	31
2.3. Hipótesis operativas	34
2.4. El diseño y el procedimiento	35
2.5. Los datos y las conclusiones	38
2.6. Contraste estadístico de hipótesis	39
3. Metodologías descriptivas I: encuestas	41
3.1. Presentación de un caso	41
3.2. Población, muestra y unidad muestral; representatividad y tamaño de la muestra	44
3.3. La selección de las unidades muestrales. Tipos de muestreo probabilístico. Muestreo por conglomerados y etapas	46
3.4. Muestreos no probabilísticos: accidental y a propósito	49
3.5. Elaboración del cuestionario	51
3.6. El trabajo de campo y el medio a partir del cual se encuesta ..	54

Capítulo II. Metodologías descriptivas II: la observación	59
--	-----------

1. La observación al servicio de la Psicología científica	59
---	----

1.1.	Presentación de un caso	59
1.2.	Observación científica: sistemática y replicable	60
1.3.	Diferentes elementos del proceso de observación	61
1.4.	El objeto de observación: marco teórico y nivel de análisis ..	64
2.	Diferentes modos de plantear la observación	66
2.1.	Tipos de observación	66
2.2.	Tipo de código	69
3.	Construcción de códigos	70
3.1.	Pasos en la elaboración de un código	70
3.2.	La pregunta, el nivel de análisis y la observación previa	71
3.3.	El primer conjunto de categorías exhaustivas, excluyentes y con suficiente nivel de detalle. Su depuración	75
4.	Tipos de medida en la observación	77
4.1.	Medidas: ocurrencia, frecuencia, latencia, intensidad, duración	77
5.	La evaluación del proceso de observación	81
5.1.	Fiabilidad: acuerdo entre observadores	82
5.2.	Validez: de contenido, de constructo y referida al criterio ...	83

Capítulo III. Metodología experimental 89

1.	Experimentos con una variable independiente; diferentes sujetos ..	89
1.1.	Introducción: ¿qué es experimentar en Psicología? Presentación de un caso	89
1.2.	Diseño con grupos aleatorios: la hipótesis; los participantes y su asignación; el diseño, las variables y su operativización; los resultados y la discusión	92
2.	Experimentos con una variable independiente; los mismos sujetos .	103
2.1.	Presentación de un caso	103
2.2.	Ventajas de experimentar con los mismos sujetos: control y economía. Inconvenientes: aprendizaje, fatiga, motivación y práctica	106
2.3.	Tipos de diseños para controlar el efecto de la práctica: aleatorización simple; aleatorización en bloques; reequilibrado y cuadrado latino	109
3.	Experimentos con un solo sujeto	113
3.1.	Presentación de un caso. El control; la intervención y la retirada. Tipos de diseños con un solo sujeto	113

4. Experimentos con más de una variable independiente: diseños factoriales	117
4.1. Presentación de un caso	117
4.2. Los elementos de un diseño factorial: la representación gráfica, los efectos simples, la interacción y los efectos principales ...	126
4.3. La interpretación de los resultados factoriales. Generalización a más variables y a los cuasi experimentos	129

Capítulo IV. Metodologías cuasi experimentales y selectivas 135

1. Metodologías cuasi experimentales	135
1.1. Presentación de un caso: introducción al prefijo <i>cuasi</i>	135
1.2. Diseños de medidas antes y después del tratamiento. El grupo de cuasi control	139
1.3. Otros modos de mejorar el diseño pre-post: otra cohorte, otra variable	142
1.4. Los diseños de serie temporal interrumpida	145
1.5. La validez de los diseños cuasi experimentales	148
2. Diseños ex post facto	151
2.1. Una historia real. Un caso de investigación <i>ex post facto</i>	151
2.2. Estrategias retrospectivas	153
2.3. Estrategias prospectivas	158
3. Diseños evolutivos	162
3.1. La edad como variable independiente	162
3.2. Diseños evolutivos transversales	163
3.3. Diseños evolutivos longitudinales	164
3.4. Diseños evolutivos secuenciales	166
4. Resolución de casos prácticos	167

Capítulo V. Anexo 173

1. Algunas controversias metodológicas	173
1.1. Las metodologías cualitativas como modo alternativo de investigación	173
1.2. El renacimiento subjetivista	177
2. La estructura del informe científico: las normas de publicación APA '01	179
3. La ética en la investigación. El código deontológico del psicólogo ..	183

Bibliografía	186
Glosario	192

Prólogo

"Sobre la Psicología y sus métodos". Éste podría ser un subtítulo de la obra que aquí introducimos. Puede observar que ponemos primero *Psicología* y después *métodos*. Tiene que ser así. El método es una herramienta para el conocimiento psicológico; por tanto, primero el fin, después el medio. Además, expresa la filosofía con que está escrita esta obra: se trata de una metodología encarnada; no de una metodología en abstracto, ni de una metodología para la metodología. Hablamos de la manera en que se hace Psicología. Cada técnica aparece en un contexto, no como un ejemplo en el que se muestra cómo se aplicaría, sino que se da una situación que necesitará una cierta técnica para resolverla. Después vendrán las definiciones (si hacen falta). Por tanto, dispóngase a leer Psicología y, de paso, aprenderá algo de la metodología con que se ha construido.

Esta obra ha sido diseñada y ejecutada para ser trabajada con ayuda de la página web. La hemos escrito pensando que está ahí, cerca de nosotros. Hemos considerado que, además de leer y sacar notas, le gustaría, de vez en cuando, hacer alguna actividad. Por eso, tendrá la oportunidad de hacer de observador y registrar unas conductas grabadas en vídeo, podrá sortear un grupo de participantes para llevar a cabo una investigación, podrá participar en una investigación en directo, podrá leer casos diferentes de los ejemplos propuestos, tendrá un artículo de investigación -íntegro- para que lo consulte, etc. Esperamos que aprenda disfrutando.

No crea que la ausencia de profesor en el momento del primer aprendizaje le evitará aquella situación agri dulce de ser preguntado y de responder bien. Con la ayuda de la página web, cada cierto tiempo, hemos preparado preguntas de comprensión de lo que se expone en el capítulo. Podrá comprobar, con la elección de la alternativa correcta, que efectivamente lo ha entendido bien. Si no ha memorizado algún concepto y necesita conocerlo para entender un nuevo pasaje, no se preocupe: todos los conceptos tienen su explicación activable en la pantalla. Sólo tiene que hacer clic sobre la palabra en hipertexto y allí aparecerá la definición. Si quiere saber más cosas sobre una referencia utilizada, sólo tendrá que ir al final y consultar el glosario: activando el nombre de los autores tendrá acceso a la referencia completa.

¿Le gustaría saber cómo va avanzando en el conocimiento de la obra? Es muy fácil. Al final de cada capítulo en la página web tiene un test de autoevaluación. Este test no será siempre exactamente el mismo, de manera que podrá repetirlo. Hemos pensado que a todos nos gusta empezar haciendo pruebas que no sean muy difíciles; por ello, el test tiene una dificultad variable que se ajustará a su nivel de progreso. También le informará de la dificultad que ha superado. Cuando ya tenga toda la materia preparada podrá optar por un test general de todos los contenidos.

El objetivo principal de esta obra es que el lector aprenda *cómo* se ha obtenido el conocimiento psicológico, *por qué* se ha hecho con esta metodología y cuáles son las *ventajas* y los *inconvenientes* que plantea.

Como objetivos secundarios podemos señalar los siguientes:

- Que el lector conecte el contenido de la epistemología y del análisis de datos por medio del eslabón que los une. Una determinada orientación del conocimiento debe desembocar siempre sobre un plan para obtener información de la realidad. Esta información cuando tiene forma numérica se procesa mediante el análisis de datos.
- Que el lector sea capaz por sí mismo de leer y entender los documentos científicos originales que transmiten los resultados de las investigaciones. Todas las fuentes documentales de la Psicología están comunicadas inseparablemente de la metodología empleada.
- Que el lector desarrolle un espíritu crítico respecto a la forma de acceder al conocimiento. No sólo del saber académico, sino también de toda la información que le llega por los medios de comunicación: de las entrevistas, de las encuestas, de los programas de actuación social, de los debates sobre educación, de las secciones de divulgación científica, etc.

En definitiva, nos gustaría que el lector de esta obra incorporase la metodología a su bagaje de herramientas de pensamiento y que le sea útil en cualquier función que desarrolle en su vida.

Capítulo I

Introducción y metodología de encuestas

1. La conexión epistemológica

1.1. Presentación: un relato ilustrativo

"Una tarde de color de plomo, más triste por ser de primavera y parecer de invierno, la Regenta, incorporada en el lecho, entre murallas de almohadas, sola, oscuro ya el fondo de la alcoba, donde tomaban posturas trágicas abrigos de ella y unos pantalones que don Víctor dejara allí; sin fe en el médico, creyendo en no sabía qué mal incurable que no comprendían los doctores de Vetusta, tuvo de repente, como un amargor del cerebro, esta idea: 'Estoy sola en el mundo'. Y el mundo era plumizo, amarillento o negro, según las horas, según los días; el mundo era un rumor triste, lejano, apagado, donde había canciones de niñas monótonas, sin sentido; estrépito de ruedas que hacen temblar los cristales, rechinar las piedras, y que se pierde a lo lejos como el gruñir de las olas rencorosas; el mundo era una contradanza del sol dando vueltas más rápidas alrededor de la tierra, y esto eran los días; nada. Las gentes entraban y salían en su alcoba como en el escenario de un teatro, hablaban allí con afectado interés y pensaban en lo de fuera: su realidad era otra, aquello la máscara. Nadie amaba a nadie. Así era el mundo y ella estaba sola. Miró su cuerpo y le pareció tierra. 'Era cómplice de los otros, también se escapaba en cuanto podía; se parecía más al mundo que a ella'. 'Yo soy mi alma' dijo entre dientes, y soltando las sábanas que sus manos oprimían, resbaló en el lecho, y quedó supina mientras el muro de almohadas se desmoronaba. Lloró con los ojos cerrados. La vida volvía entre aquellas olas de lágrimas. Oyó la campana de un reloj de la casa. Era la hora de una medicina. Era aquella tarde el encargado de dársela Quintanar y no aparecía. Ana esperó. No quiso llamar, y se inclinó hacia la mesilla de noche. Sobre un libro de pasta verde estaba un vaso. Lo tomó y bebió. Entonces leyó distraída en el lomo del libro voluminoso: *Obras de Santa Teresa, I*.

Se estremeció, tuvo un terror vago; [...] ¿No se quejaba de que estaba sola, no había caído como desvanecida por la idea del abandono?... Pues allí estaban aquellas letras doradas: *Obras de Santa Teresa, I*. ¡Cuánta elocuencia en un letrado! ¡'Estás sola! Pues ¿y Dios?'

El pensamiento de Dios fue entonces como una brasa metida en el corazón; todo ardió allí dentro en piedad; y Ana, con irresistible ímpetu de fe ostensible, viva, material, fortísima, se puso de rodillas sobre el lecho, toda blanca; y ciega por el llanto, las manos juntas temblando sobre la cabeza, balbuciente, exclamó con voz de niña enferma y amorosa:

-¡Padre mío, Padre mío! ¡Señor, Señor! ¡Dios de mi alma!

Sintió escalofríos y ondas de mareo que subían al cerebro; se apoyó en el frío estuco, y cayó sin sentido sobre la colcha de damasco rojo."

L. Alas "Clarín" (1886). *La Regenta*. Madrid: Alianza, 1966.

Se preguntará qué hace un fragmento de *La Regenta* al comienzo de un libro de *Metodologías científicas en Psicología*. La primera razón que podemos aducir es que entre inventar un relato y utilizar uno ya escrito -espléndidamente, por cierto- nos parece que no hay lugar a dudas. La otra razón que postulamos -la principal, la que nos sirve para justificar el inicio de nuestra exposición con un relato- es sencilla. Lo hacemos para atraer su atención, de la forma más amena posible, hacia algunos de los temas previos a la presentación de los contenidos que vamos a ir desgranando a lo largo de esta obra.

Así, entre nosotros, ¿cree que la psicología de Ana Ozores, señora de Víctor Quintanar, por sobrenombre "la Regenta", gozaba de buena salud? Esperamos que esté de acuerdo en que no; al menos, no lo parece en este fragmento. Ahora bien, desde su silla y la nuestra, parapetadas en el siglo XXI de la cultura occidental, la respuesta a la pregunta parece evidente. Sin embargo, el modo de explicar por qué el personaje de Clarín se desmaya ha sufrido, sin duda, variaciones a lo largo del tiempo desde que se escribió la novela.

La pregunta que le hemos hecho y el contexto en que está han facilitado el tipo de conocimiento al que hemos recurrido de manera conjunta. A continuación le vamos a presentar algunas de las formas de conocimiento desarrolladas a lo largo de nuestra historia para después centrarnos en una de ellas -la Ciencia-, que es en la que se sitúa nuestra tarea de enseñarle el uso de diferentes metodologías en Psicología.

1.2. Diferentes formas de conocimiento

Los humanos hemos adquirido una serie de habilidades que nos permiten afrontar situaciones sin necesidad de haber tenido previamente experiencia directa con ellas. Podemos recurrir al conocimiento existente sobre ellas. Dicho de otro modo, la especie ha desarrollado mecanismos para almacenar información útil para todos sus miembros y para transmitirla de forma independiente de los organismos que la adquirieron o generaron. A lo largo del tiempo el modo de almacenar, organizar y transmitir el conocimiento ha ido cambiando. Han aparecido modos diversos de conocer, distintas formas de conocimiento. Vamos a presentarle tres: el sentido común, la religión y la ciencia. Y lo vamos a hacer dentro del contexto provisto por el fragmento literario con que hemos comenzado¹.

Cuando se lee la novela llama la atención la soledad en que se mueve la protagonista. Si Ana Ozores hubiera tenido una buena amiga seguramente la hubiera hecho partícipe de sus preocupaciones, sus anhelos y deseos defraudados. La hipotética amiga habría recurrido a su sentido común, a su experiencia, para aconsejar a la Regenta. Lo que se conoce como el sentido común, que algunos sostienen que es el menos común de los sentidos, hace referencia a una forma de conocimiento, no contrastado pero socialmente admitido, que refleja el destilado de experiencias compartidas dentro de un mismo grupo o sociedad. Muchos de los dichos populares -como los que se recogen en el refranero- no son más que expresión de esta forma de conocimiento².

Los padecimientos del personaje femenino de Clarín son afrontados desde otras formas de conocimiento. En la novela se describe el ámbito tradicional de la ciudad de Vetusta -trasposición literaria del Oviedo del finales del siglo XIX-. Tal descripción recoge el hecho de que en los círculos femeninos de clase media y alta, la actividad religiosa ocupa un lugar importante. Dentro de ella, el hábito de la confesión frecuente daba una tremenda importancia a la figura del padre espiritual. Ana Ozores recurre reiteradamente a Fermín de Pas -magistral de la diócesis- para entender el sentido de sus padecimientos y encontrar tranquilidad dentro de su habitual desasosiego. El personaje del confesor asume que su papel tiene que ir mucho más allá del de notario de culpas y pecados, y se siente preparado para orientar, aconsejar y dirigir la vida de su feligresa. Está convencido de que en el conocimiento religioso está la clave para todas las preocupaciones humanas, incluidas las más munda-

1. Puede consultar Mundó (2001) para un análisis epistemológico de los mecanismos individuales de conocimiento

2. Puede consultar Maldonado (1966).

nas, y por eso, entre otras cosas, recomienda lecturas piadosas como la que aparece en el fragmento. Esta forma de conocimiento tiene una larga tradición. Aunque cada cual pueda tener su propia visión de la experiencia religiosa, no cabe duda de que en las diferentes tradiciones, con sus respectivos libros sagrados, se puede encontrar un buen montón de conocimiento compartido, potencialmente útil para la resolución de muchas de las tareas vitales de los seres humanos.

La ciencia médica de la época no sale muy bien parada en la novela del autor asturiano. El pensamiento de Ana es claro al respecto: "sin fe en el médico, creyendo en no sabía qué mal incurable que no comprendían los doctores de Vetusta". En realidad, el autor quiere transmitirnos la visión de que los médicos del país están fuera de onda, porque lo cierto es que se nos presenta a la Ozores como prototipo de persona que padece una histeria de conversión. Cuando se escribe la novela, la psiquiatría europea estaba gestando al psicoanálisis freudiano como forma de terapia más adecuada para abordar los problemas médicos sin origen orgánico conocido, de los cuales la histeria de conversión es el ejemplo más claro.

A lo largo de toda la novela se nos da a entender que los problemas de salud de la protagonista -desmayos incluidos- son fruto de su infancia desgraciada y de sus deseos de maternidad frustrados por el desinterés del marido; en última instancia, de la energía afectiva desatendida. O en otras palabras, de la pulsión sexual insatisfecha. En esa situación, lo más factible -dada, también, la belleza de la joven dama- es que se vea abocada al adulterio. Ante esa posibilidad, la autoridad del confesor es la mejor ayuda para mantener la virtud. El conflicto está servido...

Esta lectura psicoanalítica que acabamos de hacer de la novela de Clarín tiene la intención de llamar su atención sobre la importancia del referente histórico y cultural del conocimiento científico. ¿Cómo pudo el autor asturiano construir así su relato si, cuando publicó la novela, Freud todavía estaba estudiando con Charcot en París? A continuación le ofrecemos algunas pistas para responder a esta pregunta a la vez que le presentamos la evolución del concepto de histeria de conversión dentro de las clasificaciones psiquiátricas contemporáneas (Tabla 1.1.).

Charcot, Freud, la histeria de conversión y *La Regenta*

La palabra *histeria* es la versión latina de la griega *ustera* (útero o matriz). Se aplicó ya en la antigüedad para designar determinados trastornos psicológicos que, aparentemente, sólo se daban en las mujeres. El término de *conversión* hacía referencia a la existencia de un trastorno funcional fisiológico para el que no se encuentra causa orgánica. Así pues, en el lenguaje de la época, la histeria de conversión es un trastorno fisiológico, sin causa orgánica apreciable que padecen las mujeres y

que se cree asociado al útero. Fue la escuela francesa de Psiquiatría, por boca de Charcot, la que en primer lugar sistematizó los trastornos psiconeurológicos conocidos en la época e incluyó la histeria de conversión -también masculina- entre ellos (Charcot, 1887). Freud utilizó su experiencia al tratamiento de casos de este trastorno mediante la hipnosis como punto de arranque para su teoría dinámica de la personalidad y la psicopatología humanas. Desde entonces el concepto ha sufrido cambios en su significado y su clasificación tal como se refleja en la Tabla 1.1.

Tabla 1.1. Evolución en el modo de denominar y clasificar la histeria de conversión de la A.P.A.

DSM (1952)	DSM-II (1968)	DSM-III (1980)	DSM-III-R (1987)	DSM-IV (1994)
TRASTORNOS PSICONEURÓ-TICOS	NEUROSIS	TRASTORNOS SOMATOFORMES	TRASTORNOS SOMATOFORMES	TRASTORNOS SOMATOFORMES
	Neurosis histéricas: conversión disociativa	Trastorno de conversión	Trastorno de conversión	Trastorno de conversión
		Dolor psicógeno	Dolor somatoforme	Trastorno del dolor
	Neurosis de despersonalización	TRASTORNO DISOCIATIVO	TRASTORNO DISOCIATIVO	TRASTORNO DISOCIATIVO

Los autores del trabajo a partir del cual hemos desarrollado la tabla (Chorot y Martínez-Narváez, 1995) señalan que los cambios más significativos se producen entre 1968 y 1980. En el DSM-II la influencia del psicoanálisis fue grande; de ahí la introducción de su terminología. Cuando dicha influencia es relegada, la neurosis histérica de conversión pasa a ser un trastorno de conversión dentro del grupo de los somatoformes. La neurosis histérica disociativa pasa a ser clasificada como trastorno disociativo. Dicha clasificación se mantiene hasta el DSM-IV.

Esto refleja lo acontecido desde Freud hasta nosotros, pero no nos aclara el lapso entre la publicación de *La Regenta* (1886) y los primeros trabajos de Freud sobre la histeria (Breuer y Freud, 1893). Siendo Clarín un hombre culto -catedrático de Economía-, es muy probable que llegara a conocer algunas de las obras contemporáneas de la psiquiatría francesa y ello le permitiera caracterizar con tanta claridad los trastornos "nerviosos" de Ana Ozores. Por otro lado, cabe pensar que el propio Freud es un hombre de su tiempo y que más que un genuino "inventor" de la teoría del inconsciente, lo que hizo fue estudiar y sistematizar algo que estaba impregnando el ambiente cultural de la época. Una última curiosidad: el citado trabajo de

Breuer y Freud (1893) sobre la histeria presenta el caso de una mujer austriaca llamada Bertha Pappenhein, pero los autores se refieren a ella con el sobrenombre de Ana O.

Le hemos presentado tres formas de conocimiento de manera explícita, pero hay algunas más. Sin ir más lejos, en lo que llevamos de texto hemos hecho referencia a la epistemología -disciplina filosófica dedicada a teorizar sobre el conocimiento, en griego, *episteme*- y hemos practicado, de forma *amateur*, la crítica literaria, en la medida en que hemos interpretado las intenciones del autor de la novela. Con todo ello, lo que hemos querido poner de manifiesto es que las diferentes formas de conocimiento pueden llegar a competir para abordar un problema -como es el caso de las tres primeras, sentido común, religión y ciencia- o discurrir por contextos independientes como en el caso de la epistemología y la crítica literaria. Lo importante es tomar conciencia de que todas ellas implican un ámbito de experiencia y unas reglas -no necesariamente explícitas- para su aplicación y desarrollo. Veamos las características propias de la que aquí nos interesa: la ciencia, o actividad científica, como se prefiera.

1.3. Características de la ciencia como actividad que genera conocimiento

La mitología, la religión, la filosofía y la ciencia se presentan a veces como una secuencia lógica de mejora en los modos de generar el conocimiento por parte de los humanos. Puede que haya sido así, al menos en lo que a la cultura occidental se refiere. Si lo traemos a colación no es tanto porque seamos defensores de tal postura, sino porque nos interesa mucho empezar nuestra exposición dejando una cosa clara: lo que hoy conocemos como ciencia tiene un origen histórico muy reciente y un desarrollo muy veloz, de tal modo que incluso ya hay quien apuesta por delimitar una nueva actividad cualitativamente diferente conocida como tecnociencia (Echevarría, 1999). Pero estamos yendo muy rápido. Vamos a empezar por el principio³.

Un par de definiciones recogidas por Mundó (2001) pueden servirnos como punto de partida. La primera, de Popper (1959), dice que la ciencia es:

3. Véase, por ejemplo, Dampier (1972).

"el conocimiento sobre el universo formulado mediante principios explicativos sostenidos por la observación empírica, y sujetos a la posibilidad de refutación empírica".

La segunda, de Simpson (1964), sostiene que:

"la ciencia es una exploración del universo material que busca relaciones naturales y ordenadas entre los fenómenos observados y que es autocrítica".

La constatación de sus comunidades y la de sus diferencias nos ayudan a explicar *qué es esa cosa llamada ciencia*, si nos permite que copiemos el título de otro libro muy recomendable para introducirse en el tema (Chalmers, 1984).

El título de este subapartado conlleva una definición: la ciencia es una actividad que genera conocimiento. Pero es una definición inacabada. Si añadimos la primera parte de las de Popper y Simpson, podríamos decir que:

- 1) La ciencia genera conocimiento sobre el universo –o universo material–.
- 2) Dicho conocimiento se genera, bien mediante la búsqueda de principios explicativos sostenidos por la observación empírica (Popper), bien mediante el establecimiento de relaciones naturales y ordenadas entre los fenómenos observados (Simpson).

Habrá notado que en la segunda parte el acuerdo es algo menor. Aunque es adelantar acontecimientos, le advertimos que el significado de la expresión "mediante principios explicativos sostenidos por la observación empírica" tiene connotaciones metodológicas diferentes del de "establecimiento de relaciones naturales y ordenadas entre los fenómenos observados".

La primera expresión parece abocarnos al uso de la metodología experimental que, como verá en el capítulo III, es la que permite contrastar esos "principios explicativos" que se formulan teóricamente.

La segunda es menos restrictiva metodológicamente y permite hacer uso de más estrategias metodológicas –como las recogidas en el resto de los capítulos– dentro del paraguas de la definición de ciencia.

Pero la mayor diferencia se encuentra entre la expresión *sometidas a refutación empírica* y la expresión *exploración autocrítica*. Ambas versan sobre los procesos de admisión de nuevos elementos al bagaje del conocimiento ya establecido. La primera prescribe que el proceso sea la *falsación*. Ésta consiste en la aplicación sistemática de la lógica implícita en la regla del *modus tollendo tollens*. No hace falta que abramos un paréntesis para darle una clase de lógica matemática. Basta con que entienda que el objetivo de esa regla es rechazar

–refutar, falsar– un postulado del que se desprende una consecuencia observable que no se cumple. Buscar los casos en que el postulado se cumple para apoyarlo es la estrategia contraria, la *verificación*. Ésta se basa en otra regla lógica, el *modus ponendo ponens*.

Un ejemplo de verificación sería postular que la histeria de conversión está causada por una fuerte represión sexual, aduciendo que se conocen varios casos en los que es así. La estrategia falsacionista llevaría a buscar casos en los que aparece la consecuencia, la histeria de conversión, pero no la causa, la represión sexual. Encontrar casos de histeria sin represión nos llevaría a rechazar el postulado que las relaciona, al menos como ley universal.

Una de las grandes discusiones de los filósofos de la ciencia ha sido si la estrategia lógica que seguía –o debería seguir– la ciencia es la verificación o la falsación. El círculo de Viena defendió la primera actitud mientras que Karl Popper fue el genuino defensor de la segunda. En la actualidad se pueden manejar otras que no impliquen una lógica formalizable o prescriptiva, como tendremos ocasión de comprobar en el último apartado de esta primera parte del capítulo⁴.

Baste aquí señalar que la segunda definición, la de Simpson, al ser más laxa, permite recoger esas otras visiones no logicistas sobre el proceso de revisión de la calidad del conocimiento científico. Estas visiones, en cualquier caso, también implican un procedimiento de autorregulación desde dentro de la propia actividad científica.

En la actividad 1.1 de la página web, tendrá ocasión de comprobar que nuestra forma espontánea de razonar tiene peculiaridades encasillables dentro de alguno de estos dos tipos de estrategias lógicas que acabamos de mencionar: el verificacionismo y el falsacionismo.

En la página web se dispone de una actividad complementaria (capítulo I, apartado 1.3).

1.4. Cómo se genera y organiza el conocimiento científico

Cada una de las dos formas lógicas de razonar lleva dentro de sí una forma metodológica de proceder. Así, el verificacionismo se asocia al *método inductivo*, mientras que el falsacionismo está estrechamente ligado al *método deductivo*.

4. Véase Campos y otros (2001).

El método inductivo parte de la observación cuidadosa de los fenómenos que se estudian para tratar de descubrir regularidades en su dinámica, regularidades que acaban postulándose como leyes generales del funcionamiento de la naturaleza. El método inductivo podemos caracterizarlo como un método que genera conocimiento desde abajo hacia arriba.

En la historia de nuestra cultura se señala a Aristóteles como el primer gran defensor de este método. Este gran filósofo de la Antigüedad trataba, de algún modo, de contrarrestar las ideas de su maestro, Platón, que fue el gran defensor del otro método, el deductivo. Platón, convencido de que los sentidos nos dan información engañosa a la hora de alcanzar el conocimiento de las cosas –en realidad, de las ideas–, defendía a la razón como única capacidad para dicho conocimiento. Mediante la dialéctica se conocía la idea general. Una vez identificada ésta por el razonamiento dialéctico, se comprobaba su veracidad al aplicar la idea a nuevos ámbitos de la experiencia.

El método deductivo, por tanto, parte de la formulación general para aplicar el conocimiento al caso particular. Se puede caracterizar como un método que genera conocimiento de arriba abajo.

A lo largo de la historia de la cultura occidental, diferentes filósofos y científicos han discutido largamente sobre la idoneidad de uno u otro método para el progreso de la Ciencia. Hoy por hoy nadie discute que ambas formas de generar conocimiento están fuertemente implicadas en el trabajo de los científicos.

Al método que combina ambas maneras de trabajar se le conoce como método hipotético–deductivo.

Más adelante lo detallaremos más al hilo de la aplicación en el ámbito de la Psicología. Un ejemplo de la interacción entre ambos tipos de estrategias, la inducción y la deducción, puede apreciarse en el siguiente caso, que versa sobre un reciente descubrimiento astronómico.

Observaciones que desbordan las teorías

En la sección de "Sociedad" del diario *El País* del 17 de mayo de 2001 aparecía una noticia a toda página: "Descubierta una gigantesca burbuja de agua alrededor de una estrella que nace". La noticia hacía referencia a la observación de una burbuja de agua envolviendo a una estrella en formación de un tamaño aproximado de una vez y media el sistema solar. En el cuerpo de la noticia se recogían palabras del director del proyecto, José María Torrelles, investigador del CSIC en el Instituto de Estudios Espaciales de Cataluña. "Es la primera vez que un fenómeno de estas características es observado, una estrella joven expulsando una estructura esférica. Lo novedoso, lo que ha sorprendido a los teóricos, es que no hay modelos para explicarlo". Es decir, desde el conocimiento actual no hay modo de explicarse por qué se ha formado semejante burbuja. No podría haberse deducido que algo así podría

pasar. La observación de este nuevo fenómeno obligará a cambiar algunos de los conocimientos asentados en el terreno de la astronomía, obligará –por inducción– a reformular alguna de las teorías hasta ahora imperantes.

¿Por qué no se había observado hasta ahora nada parecido? Debido a la técnica empleada por el telescopio que utilizaron. Bueno, habría que decir que lo alquilaron. Es más, habría que decir que alquilaron un conjunto de telescopios situados por todo el territorio de Estados Unidos. El conjunto se llama VLBA y pertenece a la Fundación Nacional de la Ciencia de aquel país. Consiste en un conjunto de diez radio-antenas situadas entre las islas Hawai y la costa Este pasando por las islas Vírgenes en el Caribe. Funcionando coordinadamente logran una precisión 200 veces superior a la del mejor telescopio anteriormente utilizado, el telescopio espacial *Hubble*. El periodista que redactó la noticia utilizó una imagen elocuente para que los profanos nos hagamos una idea de lo que implica semejante alcance y precisión: es como ver un céntimo a cuatro mil kilómetros de distancia⁵.

Para completar lo relativo a cómo se genera el conocimiento científico sólo necesitamos añadir una característica que no tiene nada que ver con la naturaleza de su lógica, sino más bien con la logística de su implantación.

Sea cual sea la estrategia metodológica que siga una investigación, los resultados que obtenga no serán homologados –admitidos– si no son susceptibles de ser replicados, repetidos con las mismas condiciones. Esta característica de la *replicabilidad* es central en toda la actividad científica.

Pero nos falta todavía hacer referencia al modo de organización del conocimiento científico. Lo primero que hay que señalar es que, ya sea mediante la razón, ya sea mediante la observación, el conocimiento científico se formula mediante leyes de ámbito universal o, al menos, de ámbito lo más general posible.

Dichas leyes se han formulado primero como *hipótesis*, como proposiciones tentativas para explicar un conjunto de fenómenos delimitado. Una vez que una hipótesis ha sido contrastada, pasa a considerarse una *ley* con su ámbito de generalidad definido. Las leyes se agrupan en conjuntos denominados *teorías*. Tales conjuntos sirven para dar cuenta de ámbitos cada vez más grandes de los fenómenos en exploración, poniendo en relación conocimientos de índole semejante.

Uno de los conjuntos más famosos de la Física es el de las *tres leyes del movimiento* formuladas por Isaac Newton (1642-1727) que constituyen la teoría de la mecánica clásica, teoría que mantuvo su vigencia hasta la aparición de la teoría de la relatividad a principios del siglo XX.

5. <http://www.elpais.es/diario/buscador/searcher.html>.

1.5. La ciencia de la Psicología

Cuando hemos hablado de formas de conocimiento hemos hecho referencia a que todas ellas son definidas por el ámbito de la experiencia al que se refieren y por el conjunto de reglas que establecen para su uso, su almacenaje y su ampliación. Cuando hablamos de la Psicología –una forma científica de conocimiento– debemos también definir su ámbito y sus reglas.

Desde que se considera constituida la Psicología como disciplina científica, hace algo más de 120 años, las discusiones sobre cuál es el ámbito y el método de la Psicología han sido muy numerosas y apasionadas. Haciendo un resumen muy rápido, podríamos decir que hasta la Segunda Guerra Mundial coexistieron grandes enfoques teóricos que defendían visiones muy distintas del ámbito disciplinario y del modo de investigar sobre él. El estructuralismo, el funcionalismo, el conductismo, la fenomenología, la *Gestalt*, el psicoanálisis, el marxismo... generaron otras tantas Psicologías.

A los conjuntos de teorías que rivalizan por dar cuenta del mismo ámbito de conocimiento se les denomina *paradigmas*.

Fue Tomas S. Khun en su ya clásico libro *La estructura de las revoluciones científicas* (Khun, 1962) quien acuñó dicho término para explicar el progreso de la Ciencia no como una acumulación paulatina de conocimiento, sino como fruto de una dinámica de lucha entre paradigmas rivales. En dicha lucha no sólo se recurre a argumentos lógicos, sino también a argumentos sociológicos. Es decir, no sólo se imponen las teorías que son mejores, sino las que son mejor defendidas dentro de una comunidad científica concreta⁶.

Aplicando esta visión de progreso científico a la Psicología, se puede establecer que el conductismo se apoderó de ella en la posguerra al quedar desmanteladas las condiciones materiales en que se desarrollaron las otras versiones de nuestra disciplina. Así pues, la Psicología pasó a ser la ciencia de la conducta y las ciencias sociales conductistas, las ciencias del comportamiento.

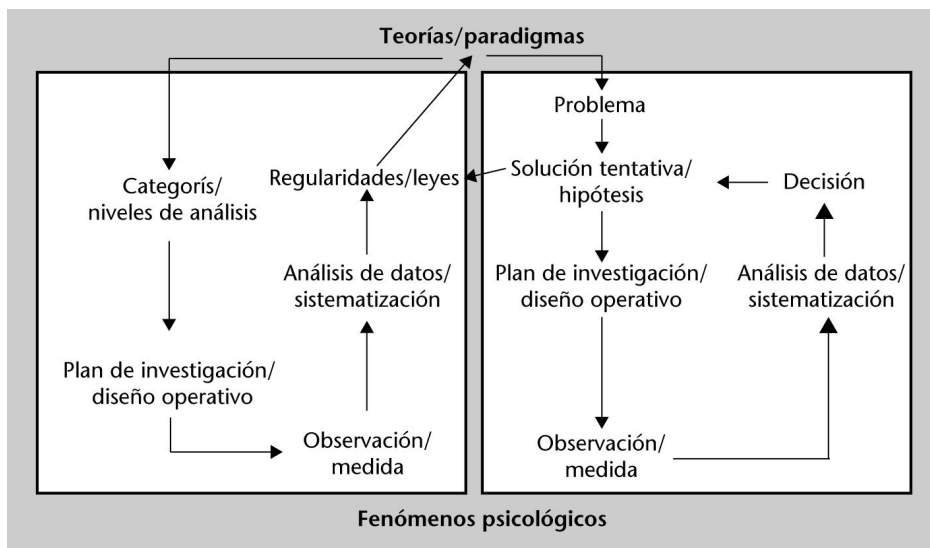
Pero, poco más de una década después de acabar la guerra, el desarrollo de los ordenadores catapultó a la Psicología cognitiva –centrada en los procesos del conocimiento humano– como alternativa al conductismo, la cual traía al corazón de la disciplina lo que aquél había postergado por mor de su opción epistemológica. En la intersección de la biología del sistema nervioso, la computación, la psicolingüística y la filosofía de la mente, la ciencia cognitiva ejerce en la actualidad como paradigma dominante⁷.

6. Véase Campos *et al* (2001).

7. Puede consultar Broncano y Mundó (2001).

En cuanto a las reglas para su investigación, podemos afirmar que son las mismas que las del resto de las disciplinas científicas si las consideramos del modo más general posible. Es decir, en Psicología utilizamos también el método hipotético-deductivo. En la Figura 1.1. se presenta un esquema del modo de proceder con el método hipotético-deductivo tal y como hoy día lo utilizamos.

Figura 1.1. El método hipotético-deductivo en psicología



Observe que en la parte de abajo hemos colocado los fenómenos psicológicos, y en la parte superior, las teorías y paradigmas que dan cuenta de ellos. Es una manera de hacerle ver que unos no se entienden sin las otras. Ahora concéntrese en la parte izquierda de la figura. En ella hemos querido representar el ciclo que tiene que ver con la inducción, la generación del conocimiento de abajo arriba. La observación cuidadosa de determinados fenómenos que nos interesan permite establecer regularidades que son postuladas como leyes. La detección de dichas regularidades se hace mediante técnicas de sistematización que, en la actualidad, son mayoritariamente estadísticas. Fíjese en que la observación está mediada por categorías de análisis que siempre tendrán una carga teórica. Las leyes así generadas se asimilan a las teorías y paradigmas ya existentes o dan lugar a nuevas teorías. En la parte derecha de la Figura 1.1 se representa la deducción. Desde el conocimiento establecido –o desde una teoría rival y alternativa– se postula una solución tentativa al problema, una hipótesis. Para ponerla a prueba se desarrolla un plan de investiga-

ción en el que dicha hipótesis se articula de forma operativa. Se llevan a cabo observaciones que dan lugar a unos datos de cuyo análisis se sigue una decisión respecto a la idoneidad de la solución –la hipótesis– planteada. Todo ello generará los ajustes necesarios en la teoría, bien para cobrar más fuerza si la hipótesis funciona, bien para ser puesta en cuestión si la hipótesis fracasa.

La obra de Freud es un ejemplo prototípico de inducción. Por los años en que se publicó la novela *La Regenta* con la que empezábamos el capítulo, la escuela francesa de Psiquiatría había constatado mediante análisis anatómicos forenses que existían una serie de trastornos fisiológicos que no tenían causa neurológica. La histeria de conversión era el ejemplo más claro. Más tarde, después de estudiar en París, Freud, junto a Breuer, utilizó la hipnosis como método para demostrar el origen puramente psicológico del trastorno. Bajo estados hipnóticos, desaparecían las dolencias fisiológicas de la histeria. El hecho de que lo ocurrido bajo hipnosis no fuera accesible al paciente una vez recuperado su estado habitual llevó al psiquiatra vienes a la formulación de su teoría del inconsciente y al desarrollo de su técnica psicoterapéutica basada en la libre asociación y el análisis de los sueños.

Como ejemplo de deducción podemos citar el surgimiento de la teoría de la motivación intrínseca. De la aplicación de las leyes del aprendizaje operante a la educación se derivaba que los niños premiados por su trabajo deberían incrementar su persistencia en las tareas reforzadas. Pues bien, Lepper, Greene y Nisbett (1973) pusieron de manifiesto que la aplicación de premios a niños que llevaban a cabo tareas libremente elegidas hacía que disminuyera su persistencia en las mismas. Este trabajo y otros similares han dado lugar a la teoría de la motivación intrínseca, teoría que rivaliza con el conductismo en la explicación de la motivación académica.

Veamos, a continuación, todo el proceso de contraste de hipótesis de un modo más detallado y ligado a un ejemplo concreto de investigación llevada a cabo en el estudio de las emociones.

2. El proceso de contraste de hipótesis

2.1. Proceso de investigación

Después de haber presentado ideas tan generales como las que nos permiten referirnos a la actividad científica y a sus peculiaridades cuando la situamos en el contexto de la Psicología, nada como un buen baño de conceptos "tangibles" para "aterrizar". A lo largo de este apartado vamos a intentar ponerle en una situación concreta que le permita hacerse una idea cabal de cómo es el *proceso de investigación* dentro de la Psicología.

Empezaremos presentando el proceso en su vertiente más deductiva:

- Éste consiste en un conjunto de pasos secuenciales que empieza por la detección y delimitación de un *problema*.
- Para resolver este problema, se postula una solución tentativa o *hipótesis*.
- De la mencionada hipótesis, se deducen consecuencias contrastables en el plan empírico, por medio del recurso a la *observación*.
- Llevar a cabo este contraste implica la elaboración de un plan, de un *diseño* de la investigación.
- Los *datos* que recopilamos por medio de este diseño nos permitirá tomar decisiones y llegar a *conclusiones* sobre la idoneidad de la hipótesis.
- Todo el proceso revertirá en la comunidad investigadora por medio de un *informe* elaborado por los autores y en el que se detallarán los pasos seguidos.

Precisamente vamos a servirnos de un informe de investigación publicado hace unos cuarenta años por Schachter y Singer (1962) –en *Psychological Review*, una de las revistas de más solera en nuestra disciplina, aparecida en 1894– en el que se aborda el problema del papel del conocimiento en la respuesta emocional, un problema clásico de la investigación psicológica. Al hilo del relato de su investigación iremos desmenuzando los pasos del proceso que hasta ahora sólo hemos esbozado.

Pero es necesario hacer una advertencia antes de empezar. El ejemplo que vamos a detallar no es más que uno de los modos posibles de investigar con una estrategia deductiva. No es el único, como tendrá ocasión de comprobar cuando lea el capítulo IV. Aquí el ejemplo ni siquiera lo mencionamos como prototipo de experimento –para eso está el capítulo III– sino como ilustración del *proceso de investigación* mediante la mencionada estrategia deductiva.

2.2. El problema

El problema de qué indicadores, internos o externos, permiten a una persona etiquetar e identificar su propio estado emocional está entre nosotros desde los tiempos en que James (1890) expuso su *doctrina*⁸ por primera vez estableciendo que 'los cambios corporales siguen directamente a la percepción del hecho emocionante, y que nuestro sentimiento de esos mismos cambios según ocurren es la emoción'. Si podemos percibir una variedad de sentimientos y estados emocionales, éstos deberán ir acompañados por una variedad de estados corporales diferenciables. Siguiendo el pronunciamiento de James, se llevaron a cabo un gran número de estudios en la búsqueda de los diferenciadores fisiológicos de las emociones." (Schachter y Singer, 1962, p. 379).

Con este párrafo empiezan su informe la pareja de investigadores norteamericanos que acabamos de citar. Fíjese con qué naturalidad definen un problema en relación con una teoría que, por entonces, tenía más de setenta años de existencia.

Ahora bien, el problema es demasiado general, tal y como está formulado en ese primer párrafo. Necesita ser traducido, al menos en parte, a unos términos tales que pueda formularse su solución de un modo operativo, contrastable.

Cuando se exponen de forma general los pasos del proceso de investigación puede dar la sensación de que una vez establecido el problema, el resto es fruto de la creatividad de investigador –en muchos casos, entendida como genialidad. Aunque no negamos que pueda haber casos que encajen con esa idea –que nos atrevemos a calificar de romántica– de la actividad investigadora, lo más habitual es que el salto entre el problema y la solución tentativa sea fruto de un concienzudo trabajo de revisión de investigaciones anteriores.

Sin negar la parte creativa del trabajo investigador, aquí nos interesa señalar que lo que más ayuda en esta fase es analizar las aportaciones previamente realizadas por otros investigadores. En los tiempos que denominamos como *sociedad de la información*, lo mejor es documentarse sobre la existencia de trabajos previos.

Volviendo al trabajo que nos sirve de ejemplo, podemos continuar el hilo argumental de los autores para ilustrar lo que acabamos de decir. ¿Cómo conectan su investigación con el problema ya planteado por William James? Me-

8. La palabra doctrina la hemos resaltado nosotros para que note su uso como sinónimo de teoría.

diante la revisión y el análisis de diferentes informes de investigación previamente publicados. Recordará que habíamos dejado el primer párrafo finalizado con la expresión:

"[...] se llevaron a cabo un gran número de estudios en la búsqueda de los diferenciadores fisiológicos de las emociones."

Pues bien, los autores nos explican a continuación que tales estudios rechazaron la idea de James de que diferentes estados fisiológicos estarían en la base de diferentes emociones. Más bien constatan la aparición de un estado fisiológico muy parecido en todo tipo de emociones. Sería algo así como una activación emocional general. Después nos cuentan que otros autores habían lanzado la hipótesis sobre:

"[...] qué factores cognitivos podrían ser los principales determinantes de los estados emocionales".

(Schachter y Singer, 1962, p. 380).

A partir de ahí se les ocurre que el producto final, la emoción, podría deberse a la acción conjunta de un estado de activación fisiológica general –inespecífica– y elementos de información relevantes para su interpretación –específica– por parte de la persona que se emociona.

En ese momento presentan la revisión de un trabajo realizado por Gregorio Marañón publicado en Francia en 1924 (Marañón, 1924/1985).

Una pista lanzada por el médico español como posible explicación de lo que ocurre en algunos casos –puede consultar el resumen extenso que le presentamos a continuación– da a los investigadores norteamericanos la idea central de su investigación. La pista consiste en la posible explicación sugerida por el médico español para algunos de sus casos.

Marañón y la adrenalina

Gregorio Marañón (1887-1960), médico y ensayista madrileño, se destacó por sus trabajos en el terreno de la endocrinología. En 1924 publica en la *Revue Française de Endocrinologie*, un trabajo titulado "Contribution à l'étude de l'action émotive de l'adrénaline", que nosotros hemos leído en su traducción castellana, realizada por nuestro compañero José Antonio Corraliza en 1985 y publicada en el número 21 de la revista *Estudios de Psicología*. Este trabajo, que fue el que resultó tan útil a Schachter y Singer, consiste en la presentación de una serie de observaciones realizadas con pacientes con trastornos endocrinológicos a los que se trataba con pequeñas dosis de adrenalina. El total de personas observadas fue de 210 y en todas ellas

se tomaron notas acerca de sus reacciones clasificadas en siete grupos: *efectos locales, efectos circulatorios, efectos respiratorios, efectos motores, efectos secretorios, efectos metabólicos y efectos subjetivos*.

Dentro del último grupo de reacciones se presenta la que el autor denomina *reacción emotiva*. En sus propias palabras:

"algunos sujetos a los que se ha inyectado adrenalina presentan [...] fenómenos emocionales que se manifiestan bajo dos formas: algunas veces *como simple percepción subjetiva de ciertos trastornos somáticos que hacen nacer en el sujeto una sensación emotiva indefinida, pero percibida "en frío", sin emoción propiamente dicha, y en otras ocasiones como una emoción involuntaria completa, es decir, con los mismos elementos somáticos que en el caso precedente y, además, con la participación psíquica afectiva que es el complemento de estos elementos*" (p. 78 de la traducción castellana).

Señala que el primer tipo es el más frecuente –no precisa cuánto– y que se caracteriza porque la persona es plenamente consciente de que no está emocionada aunque lo parezca, utilizando siempre expresiones del tipo: "siento como si tuviera miedo", "como si esperase una gran alegría", "como un sobresalto interno", "como si fuera a llorar sin saber por qué".

En cuanto al segundo tipo señala su creencia de que se produce cuando se sugiere a los enfermos algún recuerdo de gran energía afectiva que, en condiciones normales, no sería suficiente para despertar la emoción. Señala:

"por ejemplo, en varios casos hemos hablado a nuestros enfermos, antes de poner la inyección, de sus hijos enfermos o de sus padres muertos, y han reaccionado con calma a la evocación del recuerdo. Esa misma evocación, minutos más tarde, durante la conmoción adrenalínica, ha sido suficiente para desencadenar la emoción" (p. 79).

Esta idea de la conexión entre el elemento cognitivo previamente evocado y el sentimiento pleno del estado emocional una vez provocada la activación fisiológica mediante la inyección de adrenalina, proporciona a los investigadores norteamericanos, casi cuarenta años más tarde, una de las ideas sistemáticamente puestas a prueba en su estudio, es decir, una de sus hipótesis.

Estos dos estudios, con un lapso de tiempo tan largo ilustran el paso desde una primera investigación de tipo descriptivo llevada a cabo de forma sistemática en un ámbito poco estructurado, como es una consulta médica, hasta una investigación cuyo objetivo es contrastar una hipótesis causal y que, para ello, se lleva a cabo en una situación de laboratorio en el que casi todas las variables son tenidas en cuenta.

Tras la inyección de una dosis de adrenalina se produce una vívida reacción emocional cuando en la conversación anterior se había hecho referencia a acontecimientos emocionantes. En el resto de los casos o no había reacción emocional o ésta se percibía claramente como "no auténtica".

Lo que en la investigación de Marañón había ocurrido de forma espontánea puede ser sistemáticamente provocado en el laboratorio, debieron de pensar los investigadores norteamericanos.

2.3. Hipótesis operativas

"Dado un estado de activación fisiológica para el que un individuo no tiene explicación inmediata, éste etiquetará dicho estado y describirá sus sentimientos en términos de las informaciones disponibles a su alrededor".

(Schachter y Singer, 1962, p. 381).

Esta hipótesis conecta el problema de la activación fisiológica con el papel de las variables cognitivas que podrían ser las claves para la "particularización" –el etiquetado– de las emociones.

Ahora, desde la formulación inicial de un problema general, hemos pasado a formular una explicación tentativa mucho más concreta.

¿Concreta? Si lo decimos en el sentido de acotar más la aplicación de la teoría, podría aplicarse el adjetivo. Pero si queremos que la formulación de la hipótesis nos ayude a tomar una decisión clara sobre su pertinencia, tenemos que ser mucho más concretos. Tenemos que formularla como *hipótesis operativa*. Esto significa que la tenemos que delimitar en todos sus aspectos de tal modo que las observaciones que vayamos a realizar no dejen lugar a dudas. Además, al hacer operativa una hipótesis estamos facilitando que la *investigación pueda ser replicada*.

En la cita que abre este subapartado aparecen los siguientes términos: *activación fisiológica, explicación inmediata, sentimientos e informaciones disponibles*.

Todos ellos son dimensiones del problema en estudio que pueden tomar diferentes valores. A tales dimensiones se les conoce con el nombre de *variables*.

En la investigación, las cuatro mencionadas fueron hechas operativas, es decir, observables y medibles. Siguiendo la tradición de la que Marañón fue pionero:

1) La *activación fisiológica* fue concretada como la provocada por una inyección de epinefrina –término sinónimo de *adrenalina*– en una dosis de medio centímetro cúbico en una solución de 1:1000.

2) La *explicación inmediata* se concretó de formas diversas en los distintos grupos de personas que participaron como sujetos voluntarios de la investigación y consistió en la información que se les dio sobre los supuestos efectos del fármaco que se les suministraba haciéndoles creer que era una vitamina cuyo efecto terapéutico se trataba de evaluar⁹.

3) Las *informaciones disponibles* se concretaron mediante la presencia de un colaborador en la habitación donde esperaba el sujeto después de la inyección. Su actuación era un modo de suministrar información emocional de forma indirecta.

4) Finalmente, los *sentimientos* –las emociones– se midieron con las respuestas de los sujetos mediante la observación de su comportamiento mientras estaban en la sala de espera con el colaborador camuflado de participante. En el próximo capítulo profundizaremos sobre el uso de la observación.

Antes de pasar a detallar el paso siguiente del proceso, queremos presentar un tipo especial de variable que llamamos *constructo*.

Un *constructo* es una variable que no es directamente accesible a la observación. La inferimos por medio de indicadores a los que sí tenemos acceso.

Esos indicadores son definidos dentro de una teoría o, incluso, puede haber dos teorías rivalizando en la definición de los indicadores de un mismo constructo. Un ejemplo de constructo es la histeria de conversión que se "observaba" en el relato inicial del capítulo. Otros constructos clásicos de la Psicología son la inteligencia o la personalidad.

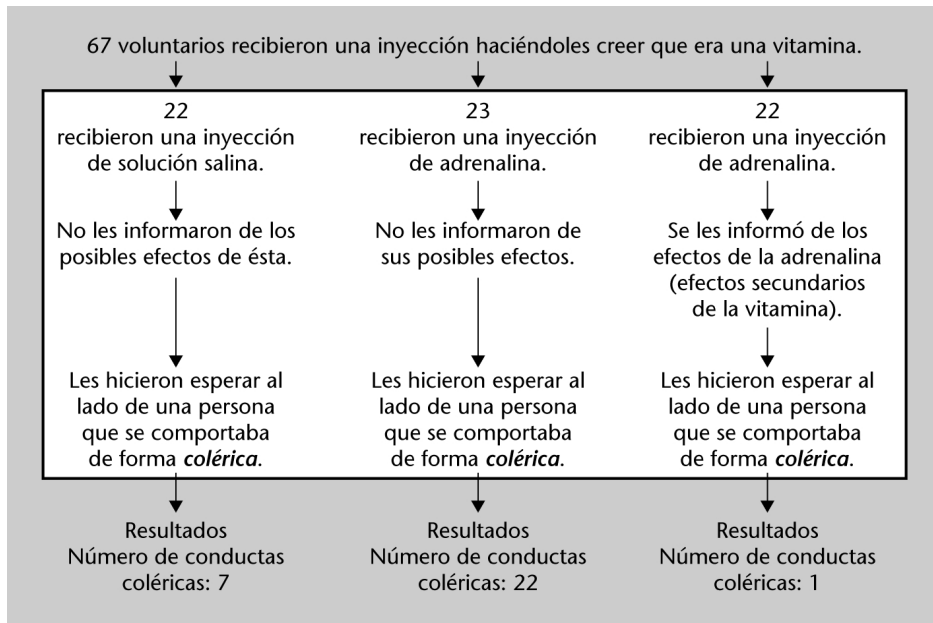
2.4. El diseño y el procedimiento

Hasta este momento del proceso, el investigador ha detectado un problema dentro de un campo del conocimiento; se ha documentado sobre los resultados obtenidos hasta el momento en el mencionado campo; ha apuntado una solución tentativa y, con todo ello, ha definido las variables implicadas y el modo de hacerlas operativas.

Ahora necesita preparar un plan de investigación –un *diseño*– que le permita poner a prueba su hipótesis en unas condiciones tales que los resultados –tanto si son favorables como si son desfavorables– sean lo más concluyentes posible. A la implantación de ese plan de investigación –tanto si es un experimento como si no– se le llama *procedimiento*.

Veamos cómo se articula este momento del proceso de investigación volviendo a la de Schachter y Singer (1962). En la figura 1.2 se presenta de forma gráfica el procedimiento que llevaron a cabo estos autores.

9. Es decir, a los sujetos participantes se les engañaba sobre el verdadero objetivo de la investigación y sobre la naturaleza real del fármaco que se les administraba. Puede consultar, en el anexo de esta obra, en qué condiciones está permitido recurrir al engaño para llevar a cabo una investigación.

Figura 1.2. Procedimiento y resultados de Schachter y Singer (1962)

A todos los participantes se les suministraba una inyección haciéndoles creer que estaban participando en una investigación sobre los efectos de un complejo vitamínico sobre la agudeza visual. Con los participantes, al azar, se formaban tres grupos.

Al primero (A) se le inyectaba solución salina de efectos inocuos.

Al segundo (B) y al tercero (C) se les inyectaba una dosis de adrenalina.

A este último grupo (C) se le informaba que la inyección podía tener determinados efectos fisiológicos secundarios transitorios –efectos que son los que produce la sustancia verdaderamente inyectada– y a los otros dos grupos (A y B) no se les decía nada.

Después, a todos se les indicaba que había que esperar unos veinte minutos a que la vitamina entrara en el flujo sanguíneo. En ese momento se les ponía en compañía de otro supuesto participante que era en realidad un ayudante de investigación. El ayudante no sabía quiénes eran del grupo informado (C) y quiénes de los otros (A y B). Tampoco sabía qué se esperaba de unos y de otros.

El compañero de espera actuaba mostrando un claro rechazo colérico de la situación. Su actuación se articulaba en una rutina que se montaba en torno a la contestación de las preguntas de un cuestionario. Las protestas empezaban con quejas relativas a la longitud del mismo, seguían con comentarios agresivos hacia la aparente estupidez de alguna de las preguntas intermedias e iban *in crescendo* hasta alcanzar el punto álgido en el momento en que el ayudante exclamaba "¡No aguanto más! ¡Me voy!", tiraba los papeles al suelo y abandonaba la sala.

La última pregunta del cuestionario rezaba así: "¿Con cuántos hombres (además de su padre) ha tenido su madre relaciones extramaritales? 4 o menos; entre 5 y 9; 10 o más".

Como ya hemos anticipado, a este plan de investigación, a este tipo de diseño en el que el investigador trata de poner a prueba una relación causal en unas condiciones en las que puede manipular y controlar las variables implicadas, se le llama *experimento*. No vamos a explicar ahora en qué consiste un experimento. A eso dedicamos todo el capítulo III. Lo que queremos es hacerlos caer en la cuenta de que el plan de investigación podría haber sido de otro tipo.

Por ejemplo, Marañón (1924), interesado en los efectos terapéuticos de la adrenalina sobre sus pacientes de endocrinología, nos informa de un estudio que consiste en la descripción detallada de efectos de esa sustancia sobre diferentes casos clínicos a los que atendía en su hospital.

Utiliza un *diseño descriptivo* mediante observación¹⁰. Teniendo ocasión de registrar tales efectos en gran cantidad de personas, recoge cuidadosamente las reacciones de sus pacientes en un conjunto de categorías previamente establecido al efecto. Ello le permite establecer regularidades. Fíjese en que, en este estudio, algunas de dichas regularidades, aun afectando a un número relativamente pequeño de casos, sirvieron para inspirar la hipótesis de un experimento posterior. Note que este tipo de investigación encaja más en un proceso de tipo inductivo (parte izquierda de la Figura 1.1), mientras que el de Schachter y Singer respondía más a un proceso de tipo deductivo (parte derecha de la Figura 1.1).

Existen diferentes planes de investigación, diferentes tipos de diseño, que se caracterizan por los objetivos que persiguen y por las condiciones en las que se llevan a cabo. Según esto, los diferentes tipos de diseño pueden clasificarse en *diseños descriptivos* (mediante encuestas u observación), *diseños cuasi experimentales* y *diseños experimentales*.

10. Véase el capítulo II de esta obra.

Ésta es la clasificación que utilizamos en este libro. Pero otros autores, incluidos nosotros mismos en anteriores ocasiones, los agrupan de modo diferente. Shaughnessy, Zechmeister y Zechmeister (2003), Kerlinger y Lee (2000), Rosnow y Rosenthal (1998), Morales (1989), Arnau, Anguera y Gómez (1990), León y Montero (2003) son ejemplos de manuales recientes, anglosajones y españoles, en los que las clasificaciones de los diseños varían con respecto a la que aquí se utiliza. No se trata tanto de que memorice variaciones sobre un mismo tema como de que tome conciencia de que la clasificación tiene un sentido didáctico más que ontológico. Creemos que al final del libro estará en mejores condiciones de opinar sobre la utilidad de nuestra clasificación de los diseños frente a otras clasificaciones alternativas. Así que, por el momento, vamos a dar por zanjado el asunto.

2.5. Los datos y las conclusiones

El final de todo proceso de investigación siempre implica volver al punto de partida. ¿Debemos rechazar, o no, la hipótesis que originó la investigación? ¿Hemos conseguido, o no, hacer una descripción sistemática de un fenómeno todavía poco explorado? El investigador sistematiza el fruto de su trabajo generando lo que se conoce como datos. Los códigos mediante los cuales se lleva a cabo esa labor de sistematización pueden ser de diferentes tipos: el lenguaje natural, los números –medidas–, gráficos, imágenes, etc.

En la historia reciente de la Psicología hay aportaciones muy relevantes que se han hecho generando datos codificados en diferentes modalidades. Las observaciones mediante las cuales Piaget analizaba el desarrollo cognitivo de sus hijos, las gráficas mediante las cuales Skinner informaba del impacto de diferentes programas de refuerzo o la cantidad de varianza explicada por el "factor g" de inteligencia en la teoría jerárquica de Spearman podrían ser ejemplos prototípicos de diferentes modos de representar los datos.

En la Figura 1.2 se presenta el comportamiento más usual dentro de cada uno de los grupos en que se dividió a los participantes.

El comportamiento colérico de los participantes se midió como número medio de indicadores de enfado. En términos numéricos, la hipótesis de los investigadores se podría expresar así:

$$\begin{array}{ccccc} \text{Grupo B} & & > & \text{Grupo C} & = & \text{Grupo A} \\ \text{mados)} & \text{(adrenalina, no informados)} & & \text{(adrenalina, informados)} & & \text{(inocuo, no infor-} \end{array}$$

Eso fue lo que los datos reflejaron (consultad la fila inferior de la Figura 1.2). La herramienta para determinar si la diferencia observada entre las medias es relevante es la estadística.

Con datos interpretamos que, una vez producida la activación mediante la inyección de adrenalina, los participantes a los que informa el investigador creen que lo que les ocurre es debido a un efecto colateral de la vitamina y no muestran emoción. Los que no han sido informados reciben indicadores para interpretar su estado en términos de la emoción que expresa el investigador camuflado. Ello permite apoyar la hipótesis de que los factores cognitivos son determinantes en la percepción de la activación fisiológica como una emoción específica.

Pero, más que discutir sobre teoría de las emociones, lo que nos interesa es que tome conciencia del modo en que los investigadores han cerrado el círculo del proceso de su estudio. Veamos ahora con detenimiento el proceso estadístico mediante el cual se decide si los datos recogidos se ajustan, o no, a lo que se esperaba en la investigación.

2.6. Contraste estadístico de hipótesis

Los investigadores recurren a la estadística para tomar esa decisión. Para poder utilizar la estadística como una herramienta mediante la que tomar decisiones sobre las hipótesis es necesario hacer previamente algunas transformaciones, algunas traducciones.

En el ejemplo que estamos siguiendo, ya se ha formulado la hipótesis de forma operativa y se ha diseñado una situación que permite su contraste. La comparación del grado en el que los participantes de las diferentes condiciones se comporten y se perciban como emocionados, tendrá que aportar elementos para tomar una decisión sobre la hipótesis.

Para hacer más sencilla la exposición, vamos a centrarnos en la parte de la hipótesis que hace referencia a las diferencias esperadas entre los grupos de participantes a los que se inyectó adrenalina pero que recibieron diferentes grados de información: los grupos B y C.

Los autores del estudio tendrán razón si, entre los inyectados con adrenalina, los sujetos informados y no informados por el investigador se comportan de forma diferente. Los investigadores estarán equivocados si ocurre lo contrario, si los dos grupos se comportan del mismo modo. ¿Cómo saber si su comportamiento es reflejo de una diferencia?

No hay forma humana de saberlo. Pero sí sabemos cómo deberían comportarse en el caso de que no hubiera tal diferencia, en el caso de que fueran iguales. Y aquí es donde entra en juego la teoría de la probabilidad. Ésta nos permite señalar el grado en el que unos datos son compatibles con la hipótesis de la "no diferencia". Ese grado se expresa en términos probabilísticos. Conociéndolo, podemos tomar una decisión. Este proceso se estudia en *Estadística Inferencial*.

La hipótesis de la "no diferencia" se denomina *hipótesis nula*. La otra, su complementaria, la que defiende el investigador, se denomina *hipótesis alternativa*.

En el ejemplo, la hipótesis nula establece que no habrá diferencias entre el grupo informado y el no informado en la puntuación de emoción obtenida mediante el cuestionario.

La hipótesis alternativa señala, por el contrario, que las puntuaciones en emoción de los dos grupos serán diferentes.

Fíjese en que este proceso de toma de decisiones mediante la estadística implica que la actividad de investigar se convierte en una búsqueda de datos que permitan rechazar hipótesis nulas para así poder mantener sus complementarias. En resumen, se busca falsar la hipótesis nula para mantener la hipótesis alternativa.

Eso es lo que hicieron Schachter y Singer. Rechazaron la hipótesis nula de que los grupos de personas informadas y no informadas se hubieran enfadado en el mismo grado (1 frente a 22).

A modo de repaso de este apartado, podemos resumir el proceso de investigación en una secuencia de pasos parecidos a los siguientes:

- 1) Detección de un *problema*. Revisión de documentación sobre el estado de la cuestión.
- 2) Formulación de una *hipótesis* o solución tentativa.
- 3) *Diseño* de una situación en la que hacer operativa una derivación observable de la hipótesis.
- 4) Sistematización de la información obtenida en forma de *datos*, preferiblemente numéricos.
- 5) Aplicación de una regla de *decisión estadística*.
- 6) Conclusión. Se deriva de las implicaciones que tiene para la investigación rechazar, o no, la hipótesis nula.

3. Metodologías descriptivas I: encuestas

3.1. Presentación de un caso

Imagine la situación siguiente: es domingo por la mañana; estamos escuchando una emisora de radio. Coincidiendo con las señales horarias el locutor dice que ahora comienza el espacio habitual para los pequeños de la casa. Hoy va a dedicar la media hora del programa a realizar una encuesta entre los oyentes para averiguar lo que piensan de las agresiones en el "insti".

Primera llamada: una señora se queja de que es intolerable la situación a que se ha llegado en las escuelas, en la que se pega a los profesores, sin que el Gobierno haga nada. El locutor recuerda a los oyentes que deben ser los niños los que llamen y se ve obligado a señalar que hoy lo que interesa son las agresiones entre escolares, excluidos los mayores.

Después de varias llamadas de niños diciendo que les parece muy mal el pegarse y que hay que respetar los derechos de los niños, llama un oyente que asegura que hace un año dos niños del curso superior le quitaron por la fuerza sus cromos de Pokémon y que su "profa" no hizo nada.

Después del tercer corte publicitario y cuando restan cinco minutos de programa, el responsable anima a que llamen niños desde otros sitios de España, ya que la mayoría de las llamadas son de la comunidad autónoma en la que está la emisora de radio.

Cuando queda un minuto el locutor hace un resumen diciendo que parece que se producen bastantes agresiones en el "insti". La mayoría las ocasionan los alumnos de cursos superiores sobre los pequeños, y, –riéndose– que parece que los protagonistas son siempre los chicos, ya que sólo ha llamado una niña diciendo que a ella la insultaban, pero no que le pegasen. Después de señalar la importancia de los valores cívicos llama al buen comportamiento de los escolares españoles.

Ahora tenemos delante el suplemento de educación de un periódico de ámbito nacional en el que se presenta un extracto de un informe hecho público por la oficina del Defensor del Pueblo (Defensor del Pueblo 2000, págs. 146-147):

"Se encuestó a 3.000 estudiantes de Educación Secundaria Obligatoria [...], la mitad chicas y la mitad chicos. Una cuarta parte de la muestra pertenecía a cada uno de los cuatro cursos de la ESO [...]"

Los participantes pertenecían a 300 centros educativos públicos y privados –fueran éstos concertados o no–, ubicados en el territorio de las distintas comunidades autónomas. [...]

El error muestral que se tuvo en cuenta [...] ha sido de +/- 2,2% (con un nivel de confianza del 95,5%).

La muestra de los alumnos se obtuvo a partir del censo de centros de Educación Secundaria Obligatoria. Se hizo de forma estratificada con afijaciones proporcionales por comunidades autónomas, tamaño de hábitat y titularidad del centro. Para el caso del género y del curso se hicieron afijaciones simples."

Respecto a la cuestión crucial, incidencia de las agresiones, el informe señala (Defensor del Pueblo 2000, p. 153):

"la categoría agresión física tiene una única manifestación que se refiere a ser *pegado por los compañeros*. [...] 4,1% en la respuesta 'a veces' y 0,7% cuando la respuesta es 'en muchos casos'."

Y más adelante (Defensor del Pueblo 2000, págs. 165-167):

"Los casos de agresión física directa son [...] en su mayoría obra de los compañeros de clase (66,7%). [...] Los cincuenta y nueve casos de acoso sexual registrados en la muestra [...] son obra de compañeros de clase en su mayoría (67,8%). [...]

En lo que se refiere a la *titularidad* del centro, únicamente se observa una diferencia significativa en el caso de la conducta *hablar mal (de un compañero)*. En los centros privados este tipo de maltrato aparece con una frecuencia significativamente mayor que en los centros públicos. [...]

El maltrato a los iguales es protagonizado mayoritariamente por los chicos. Éstos tienen un papel destacado en todos los tipos de agresiones, aunque las chicas [...] aventajan a los chicos cuando se trata de hablar mal de otros."

Con estos dos relatos le invitamos a que juegue con nosotros a encontrar el mayor número de diferencias entre las dos informaciones, la del programa de radio y la de Defensor del Pueblo. Todo lo que le parezca distinto vale, ya sea forma o contenido.

Cuando tenga su lista de diferencias compárela con la que hemos hecho nosotros en la Tabla 1.2.

Tabla 1.2. Diferencias entre las dos encuestas

Programa de radio	Informe del Defensor del Pueblo
1. Objetivo: lo que piensan de las agresiones los niños.	1. Objetivo: cómo son las agresiones en la escuela secundaria obligatoria.

2. A quién va dirigido: aunque va dirigido a los escolares, no hay forma de impedir que otras personas intervengan.	2. A quién va dirigido: a los escolares y sólo contestan ellos.
3. Ámbito: se produce una mayoría de llamadas desde la comunidad donde se hace el programa.	3. Ámbito: todo el territorio nacional.
4. Participantes: los que voluntariamente deciden llamar un domingo por la mañana a una determinada emisora de radio que estaban escuchando.	4. Participantes: parece que son elegidos por los investigadores por un procedimiento muy sofisticado, lleno de términos técnicos.
5. Contenido de la descripción: abierto, pero a expensas de las llamadas que se produzcan.	5. Contenido de la descripción: agresión física directa, acoso sexual, maltrato verbal...
6. Resultados: según las intervenciones: a. bastantes agresiones b. autores, los de cursos superiores c. protagonistas, sólo chicos	6. Resultados: a. muy pocas agresiones b. autores, los compañeros de clase c. protagonistas, tanto chicos como chicas
7. Confianza que merece la información: no se informa.	7. Confianza que merece la información: el procedimiento tiene una confianza del 95,5%.

Tanto si tiene en cuenta sus diferencias como las que hemos encontrado nosotros en la Tabla 1.2, puede llegar a la conclusión de que la encuesta hecha en el programa de radio es bastante deficiente. Podemos operativizar más este calificativo diciendo que no es científica. Alguien podría contraargumentarnos diciendo que el propósito del programa de radio no era hacer ciencia. Nosotros, a esta explicación, contestaríamos diciendo que, si los oyentes de esa emisora esperan escuchar información de la realidad escolar, entonces el resumen que hace el periodista no responde a la realidad. Si queremos tomar decisiones basadas en la realidad de la agresión en la escuela, necesitamos que la información sea ajustada.

Hay, sin embargo, algo común a los dos procedimientos: buscan describir características de un conjunto. La información sobre estas características puede obtenerse preguntando directamente a los implicados. Preguntas del estilo de "qué piensas, qué haces, cómo valoras tal cosa, qué proyectos tienes, qué producto usas, etc." son características de las investigaciones descriptivas con encuestas.

Ahora bien, estamos en una materia que busca hacer las cosas de manera científica. En el informe del Defensor del Pueblo pueden intuirse algunas de las características que hacen que una encuesta sea científica: un número de encuestados grande, un procedimiento no arbitrario de selección de los encuestados y una garantía sobre la confianza que ofrece la investigación.

3.2. Población, muestra y unidad muestral; representatividad y tamaño de la muestra

Si volvemos a nuestro caso (a partir de ahora, por nuestro caso nos referiremos siempre al informe del Defensor del Pueblo) sabemos que la población a la que iba destinada la encuesta era el conjunto de los escolares españoles de Secundaria Obligatoria. Podría haber incluido también a los padres, o a los alumnos de Bachillerato, por ejemplo.

La primera característica de una *encuesta científica* es que especifica el alcance del conjunto que va a estudiar. Esto determina dos cosas: quiénes pueden ser encuestados y cuál es el alcance de las generalizaciones.

Si quería saberse la realidad de todos los escolares españoles de Secundaria Obligatoria, ¿por qué no se les preguntó a todos? La respuesta es que físicamente no se podía (ese año eran 1.800.000, aproximadamente). ¿A quién se preguntó? Se encuestó a 3.000 de ellos. Si se hubiera podido preguntar a todos, ésa habría sido la mejor forma de conocer la respuesta.

Cuando no puede accederse a todos los elementos de una población, determinamos un subconjunto de ella, al que llamamos *muestra*.

Para que un subconjunto de elementos de una población pueda constituir la muestra de un procedimiento científico, ese subconjunto debe tener una sola propiedad: debe ser *representativa*. Una muestra es representativa cuando podemos generalizar a la población, con una cierta *confianza* los datos obtenidos en ella. Como adivinará, la clave de este negocio estriba en obtener una muestra que sea representativa.

¿Cómo era la muestra? En primer lugar podemos decir que era bastante grande, 3.000 escolares. Esto refleja el hecho de que para ser representativa una muestra debe tener un tamaño adecuado. La determinación del tamaño tiene dos partes: una difícil (en función de variables estadísticas) y una fácil (existen tablas para saber el tamaño).

Tabla 1.3. Relación de tamaños de la población, precisión y tamaño de la muestra.

Población	Precisión: error asumido					
	± 1%	± 2%	± 3%	± 4%	± 5%	± 10%
500					222	83
1.000				385	286	91
2.500		1.250	769	500	345	96
5.000		1.667	909	556	370	98
10.000	5.000	2.000	1.000	588	385	99

25.000	7.143	2.273	1.064	610	394	100
50.000	8.333	2.381	1.087	617	397	100
100.000	9.091	2.439	1.099	621	398	100
infinito	10.000	2.500	1.111	625	400	100

Los tamaños muestrales son los números del interior de la tabla. Estos tamaños están calculados para cuando se mide en porcentajes y con un nivel de confianza del 95% (tabla tomada de Arkin y Colton, 1962).

En la página web se dispone de una actividad complementaria (capítulo I, apartado 4.2).

La *precisión* se refiere al error que se asume al hacer las *estimaciones* de los valores poblacionales.

Así, en nuestro caso, cuando se dice que un 4,5% de los escolares dicen que alguna vez han sido agredidos por algún compañero, debemos entender que el valor en la población se estima que sea un porcentaje entre 2,3% y 6,7%. Esto es así porque al valor medido en la muestra (4,5%) le sumamos el error muestral, 2,2%, y se lo restamos.

Si observa la Tabla 1.3, para un mismo valor población, a medida que pedimos una precisión mayor, mayor es el tamaño necesario. Por ejemplo, para una población de 100.000, si deseamos una precisión del 5% necesitamos un tamaño de 398, pero si queremos que la precisión sea del 2% el tamaño se dispara a 2.439.

También, cuando el tamaño de la población crece, el tamaño de la muestra crece. Por ejemplo, para una precisión del 5%, cuando la población tiene un tamaño de 1.000 la muestra debe tener 286 unidades; cuando la población es 10.000 la muestra debe ser 385; como puede observar, el tamaño de la muestra no crece proporcionalmente al tamaño de la población. Recuerde que en nuestro caso se representaba (para una precisión alta, 2,2%) con un tamaño de 3.000 a una población de 1.800.000.

Por último, la confianza se refiere a la seguridad del procedimiento de estimación. Por ejemplo, una confianza del 95% significa que si repetimos la encuesta 100 veces (obteniendo cada vez una muestra aleatoria de 3.000), en 95 de ellas habremos hecho estimaciones que contienen el valor de la población. Como en los dos casos anteriores, cuanto mayor sea nuestra exigencia de confianza, mayor deberá ser el tamaño de la muestra. Esto no podemos verlo en la Tabla 1.3, porque todos los valores son para un mismo valor de confianza del 95%.

3.3. La selección de las unidades muestrales. Tipos de muestreo probabilístico. Muestreo por conglomerados y etapas

Hasta ahora sabemos que los investigadores de nuestro caso, para trabajar con una muestra representativa, en cuanto al tamaño, debían encuestar a 3.000 escolares. ¿A qué tres mil escolares? ¿A tres mil de su ciudad? Seguro que se le ocurre que si la muestra tiene que representar a toda España los encuestados deben provenir de toda España. A la ciencia que determina a qué 3.000 de los 1.800.000 se tiene que encuestar se le llama *muestreo*.

El muestreo tiene en cuenta:

- Las medidas de muestra y población.
- El tipo de variable que se medirá.
- El tipo de estadístico con que se medirá la variable.
- Las características de la población que queremos representar.

En principio, para que el muestreo aporte representatividad debe asegurar que todos los elementos de la población tienen la misma probabilidad de pertenecer a la muestra.

En nuestro caso eso habría supuesto que se confeccionara una lista con todos los estudiantes, cada uno con su número de orden alfabético y, después, que se sorteara hasta disponer de 3.000 nombres. Esto, que hace sólo unos años era impensable, ahora es relativamente fácil de hacer con un ordenador. Si las entrevistas se fueran a hacer por teléfono, éste podría ser un buen método: *muestreo aleatorio simple*. Sin embargo, si queremos que un entrevistador esté presente en cada uno de los lugares de residencia de los niños y que ayude a cumplimentar el cuestionario, el método es costoso (eso sí, aprendería mucha geografía de España).

Cuando no se utiliza el método aleatorio simple, lo que se hace es estudiar primero qué características de la población son relevantes para la investigación. Por ejemplo, en nuestro caso, decidieron representar: "comunidad autónoma, tamaño de hábitat y titularidad del centro" por un lado y "género y curso" por otro.

De las primeras características nos dicen que se representaron mediante *afijación proporcional*. Esto significa lo siguiente:

1) "Comunidad autónoma". ¿Cuántas comunidades autónomas hay, 17? Entonces en la muestra debe haber estudiantes de las 17 comunidades autónomas. ¿Cuántos de cada una? Pues un número que mantenga la misma proporción que en la población. Por lo tanto, si, por ejemplo, los escolares de la ESO de Ca-

taluña son el 18%, en la muestra de 3.000 debe haber un 18% de estudiantes catalanes (540). Con el mismo razonamiento se determina cuántos debe haber de cada una de las otras comunidades.

2) "Hábitat". Este aspecto señala el tamaño de la población donde habitan los individuos. Lógicamente, hay que establecer categorías, ya que el rango de tamaños es muy extenso. Un ejemplo de categorías puede ser: "más de un millón", de "500.000 a 1.000.000", de "250.000 a 500.000", y así sucesivamente. El paso siguiente es conocer cuántos estudiantes de secundaria hay viviendo en poblaciones con los tamaños determinados (a estos datos puede accederse por medio del INE o las consejerías de educación). Si, por ejemplo, resultase que el 40% de los estudiantes de secundaria está en poblaciones de "500.000 a 1.000.000", entonces nuestra muestra de 3.000 deberá comprender un 40% que provenga de poblaciones con este tamaño de hábitat.

3) "Titularidad del centro". Supongamos que el 66% de los centros en Cataluña son de titularidad pública, entonces, de los 540 estudiantes, el 66% deberán estudiar en centros estatales (360).

Las características siguientes nos dicen que se representaron mediante *afijación simple*. Esto significa lo siguiente:

4) "Género". La mitad chicos y la mitad chicas. Eso significa que, en el caso de los componentes catalanes, 270 fueron chicos y 270, chicas.

5) "Curso". Como hay cuatro cursos en la ESO, los tres mil estudiantes españoles deberían estar repartidos en cuatro bloques iguales: 750 de cada curso.

Lógicamente, se hacen afijaciones simples cuando las particiones de las variables son aproximadamente iguales.

Estas cinco variables: "comunidad", "hábitat", "centro", "género" y "curso" son subconjuntos definidos en la población. Cada uno de ellos divide la población en partes exclusivas. Esto quiere decir que, cuando dividimos la población entre chicos y chicas, todos los elementos quedan incluidos en un subconjunto u otro. El mismo razonamiento podemos aplicarlo al resto de las variables.

Estos subconjuntos homogéneos en los que puede dividirse la población reciben el nombre de *estratos*. Este nombre técnico es fácil de asociar con el concepto de estrato aplicado en los terrenos: cada una de las franjas homogéneas en que puede dividirse una sección de terreno. Así, en la terminología de las encuestas se dice que se han representado los estratos correspondientes a la "clase social", al "hábitat", al "género" o a la "edad".

Recapitulamos lo que tenemos hasta ahora: se decidió representar los estratos de cinco variables relevantes en el estudio de la agresión en la escuela; tres de ellas se representaron mediante afijación proporcional y los de las otras

dos mediante afijación simple. El cruce de estos estratos determina las cualidades por las que tenemos que seleccionar las unidades muestrales y los porcentajes correspondientes.

De forma más concreta: si suponemos que el 30% de los escolares catalanes de la ESO vive en núcleos de hábitat de más de un millón, entonces la muestra de los 540 estudiantes catalanes deberá tener un 30% de estudiantes de centros de Barcelona –capital–, de los cuales, el 33% deben estar en centros privados, que la cuarta parte deben ser de 3º de ESO y que de éstos la mitad tienen que ser chicas. Bien, ¿dónde están físicamente estos alumnos? Dentro de sus clases ¿no? La manera de acceder a ellos no suele ser con una lista exhaustiva de todos los alumnos, sino por un procedimiento de selección que tiene en cuenta el hecho de que se hallan agrupados.

Estas agrupaciones son las siguientes:

- 1) Primera agrupación: el municipio.
- 2) Segunda: el barrio.
- 3) Tercera: el colegio.
- 4) Cuarta: la clase.

Estas agrupaciones se denominan *conglomerados*. Por tanto, se trata de subconjuntos de unidades muestrales agrupadas de manera natural (Pulido, 1971).

Observe que un conglomerado incluye unidades heterogéneas. Los alumnos que cumplen la condición de pertenecer a un determinado conglomerado, como el municipio, van a distintos colegios, a distintos cursos, etc. Lo mismo podemos razonar respecto al conglomerado "barrio", o incluso la "clase", donde hay chicos y chicas. Aprovechando la agrupación de las unidades muestrales en conglomerados, lo que se selecciona son éstos, en vez de seleccionar directamente las unidades. Es un subterfugio para facilitar el proceso. En nuestro caso, extraeremos al azar municipios, barrios, colegios y clases.

Al proceso de descender seleccionando conglomerados hasta encontrar la unidad muestral se le llama *muestreo por conglomerados y etapas*. Y a la fase final, desde el último conglomerado hasta la unidad muestral se le llama *ruta aleatoria*.

En general, en el proceso de muestreo, el descenso en el nivel de detalle se hace por medio de los conglomerados, del municipio y sección electoral –barrio–. La ruta aleatoria comienza con la selección de la calle, sigue con la manzana, y termina en un número concreto, una planta y una letra. En el caso de nuestro ejemplo, el muestreo se desarrollaría en las siguientes etapas:

- 1) Primera etapa: sorteamos a qué municipio vamos.
- 2) Segunda etapa: dentro del municipio, sorteamos a qué barrio se va a encuestar.
- 3) Tercera etapa: dentro del barrio obtenido, sorteamos a qué colegio.
- 4) Cuarta etapa: una vez seleccionado el colegio, sorteamos la clase (A, B, C, etc.).
- 5) Quinta etapa: finalmente, con la lista de alumnos de la clase seleccionada –último conglomerado– sorteamos a cuáles de las chicas (y de los chicos) se les hace la entrevista.

Todos los sorteos de los conglomerados tienen que ser estrictamente aleatorios (más adelante, en la Actividad 3.1, se explica cómo hacer una selección aleatoria). Todo este proceso, aparentemente tedioso, se prepara de antemano y se elabora una estrategia para organizar las encuestas que correspondan a los municipios seleccionados.

El *muestreo probabilístico* es la garantía de que la muestra representa adecuadamente a la población.

3.4. Muestreos no probabilísticos: accidental y a propósito

Si el muestreo probabilístico es la garantía de representación, ¿qué ocurre con el muestreo de los que escuchaban el programa de radio? A este muestreo, condicionado al accidente de estar escuchando en ese momento esa emisora de radio, se le llama *accidental*.

El muestreo accidental es siempre *sesgado*.

Aunque se dispusiera de 10.000 llamadas (como ocurre a veces con algún programa de televisión), el muestreo es sesgado: erróneo. Para el caso de las llamadas hechas desde el programa de televisión, es sesgado porque no representa bien a la población: no representa a quien no está viendo la televisión, no representa a quien no le gusta ese programa, no representa a quien no quiere llamar por teléfono, etc.

Un tamaño grande sin representación aleatoria no sirve para nada. O, bueno, puede servir si se acota bien qué es lo que se pretende, como el caso que comentamos a continuación.

Diez años de Área Reservada

Manolo Fernández, a la sazón presentador/director de un programa de radio, decide preparar la edición de un compacto en el sello RNE-Música para conmemorar los diez años de su programa. En vez de elegir los temas que a él le parecen más representativos decide pedir a sus radioescuchas que voten los temas que más les han gustado en estos diez años de Área Reservada.

El objetivo es que la selección de temas sea la que más gusta a los que escuchan el tipo de música que él programa. Supongamos que para que participe el máximo número de gente dispone de un número 900, que puede accederse por correo electrónico, que se "refuerza" a los que llaman sorteando premios, etc. Con estas circunstancias, puede decir que los temas elegidos representan bastante bien los gustos de los seguidores del programa. Aun así, un metodólogo estajanovista puede hacerle algunas puntualizaciones.

Tiene que ir con cuidado al decir que el resultado de la elección representa las preferencias de los aficionados a la música jazzy-fusión-soul-funky, por lo siguiente:

- a) Puede haber aficionados que no sepan de la existencia de este programa.
- b) Puede haber amantes de este tipo de música que vivan en sitios donde no llegue la señal de Radio 3.
- c) Puede haber aficionados que no puedan escuchar la radio a aquella hora.
- d) Puede haber aficionados que sintonicen otro programa, porque no les gusta el presentador.
- e) Puede haber algunos que no puedan marcar números 900 desde su despacho.
- f) En penúltimo lugar, a algunos no les apetece llamar a las emisoras de radio.
- g) (Ponga otra razón usted.)

El problema de la representación no es un problema de falta de democracia. El problema es que no podemos asegurar que los que no están representados piensen lo mismo que los que sí lo están. Es posible que aquellos a los que no les apetece llamar a las emisoras de radio tengan un gusto distinto del que tienen aquellos a los que no les importa llamar.

Como ve, conseguir una buena representación no es una tarea baladí. (Aunque, en nuestra opinión, el CD conmemorativo ha quedado bastante bien.)

Existe un tipo de muestreo no aleatorio que puede ser suficientemente representativo en algunas circunstancias. Se puede recurrir a un muestreo rápido y económico, si estamos dispuestos a sacrificar un poco de fidelidad de la representación y la correspondiente información sobre los errores asumidos.

Este muestreo, llamado *a propósito* u *opinático*, se basa en el conocimiento de un experto sobre las características de la población, ubicación de las unidades muestrales y forma de acceder a ellas.

La capacidad de representación descansa en la experiencia y buen juicio del experto. Puede leer la siguiente situación donde se justificaría este tipo de muestreo.

Al hilo de la noticia

Supongamos que tenemos que hacer una encuesta de urgencia para saber cómo responden los estudiantes de segundo de Bachillerato al proyecto de ley del Ministerio de cambiar el actual sistema de un examen de selectividad por dos exámenes de selectividad. La noticia saltó ayer por la noche (19-04-01) a los medios y queremos hacer la encuesta en el transcurso de la mañana para informar en los noticiarios del mediodía. Acudimos a un experto en educación, al cual le pedimos que nos diga a qué centros de enseñanza concretos podemos ir para tener una buena representación en nuestra encuesta de urgencia.

Este experto nos dice lo siguiente: un centro de un pueblo con alumnos de entorno rural, otro de un barrio de clase social media en una ciudad de tamaño medio, otro en un barrio obrero de una gran ciudad, los tres de enseñanza pública; por último, un centro de enseñanza privada de una gran ciudad. Al extrañarnos de que no tenga en cuenta las diferencias por comunidades autónomas, nos responde que por anteriores encuestas se sabe que los alumnos de segundo de Bachillerato no se diferencian en sus opiniones respecto a la selectividad.

Con esta información se puede ir a cuatro centros que cumplan estas características, seleccionar –ahora sí– al azar una de las clases de segundo de bachiller de cada centro, y entrevistar al mismo número de chicos y chicas. Con las respuestas de un total de ochenta alumnos, y los criterios de representatividad hechos a propósito, podemos "avanzar" qué opinan los estudiantes del doble examen propuesto por el Ministerio.

3.5. Elaboración del cuestionario

El *cuestionario* es el conjunto de preguntas con el que se quiere obtener la información necesaria para el objetivo de la investigación.

En el programa radiofónico con un "qué piensas de..." o un "cuéntenos su experiencia con..." suele ser suficiente para que la persona que llama diga lo que le parezca del tema.

El problema fundamental de una sola pregunta abierta es que puede que no aparezcan los datos que buscamos y pueden darse respuestas que no necesitamos. Cuando se hace una encuesta, se elabora de manera muy pormenorizada un conjunto de preguntas que cubra todos los aspectos relevantes del objetivo. Aunque en la mayoría de los casos todas las preguntas son cerradas, a veces se incluye alguna pregunta abierta para rastrear nuevos campos. En la Tabla 1.4 presentamos el comienzo del cuestionario del Defensor del Pueblo.

Tabla 1.4. Primera pregunta del cuestionario del Defensor del Pueblo

P.1. A continuación aparecen una serie de situaciones que pueden estar sucediéndole en tu centro a algún compañero o compañera, que no seas tú, de forma continua desde que empezó el curso. Rodea con un círculo una respuesta en cada línea.				
	Nunca lo he visto en mi centro	A veces lo he visto en mi centro	A menudo lo he visto en mi centro	Siempre lo he visto en mi centro
Ignorarlo (pasar de él o hacerle el vacío)	1	2	3	4
No dejarle participar	1	2	3	4
Insultarlo	1	2	3	4
Ponerle moteles que lo ridiculizan	1	2	3	4
Hablar mal de él o ella	1	2	3	4
Esconderle cosas	1	2	3	4
Romperle cosas	1	2	3	4
Robarle cosas	1	2	3	4
Pegarlo	1	2	3	4
Amenazarlo sólo para meterle miedo	1	2	3	4
Obligarle a hacer cosas que no quiere con amenazas (traer dinero, hacerle	1	2	3	4

las tareas, pedirle las zapatillas, etc.)				
Amenazarlo con armas (palos, navajas, etc.)	1	2	3	4

Si observamos la pregunta de la Tabla 1.4 podemos apreciar varias cosas:

- a) Se detecta un trabajo previo que ha hecho posible concretar todas las situaciones que los investigadores engloban bajo el término genérico de *agresión*.
- b) Las respuestas que pueden dar los encuestados son cerradas.
- c) Las respuestas implican una graduación de la frecuencia con que se han observado en el centro.
- d) Las respuestas llevan unos códigos numéricos (1, 2, 3, 4).

Toda encuesta conlleva un trabajo previo en el cual se estudian investigaciones de otros autores. También es necesario hacer un estudio empírico piloto para recoger las experiencias de los escolares. El hecho de que las respuestas sean cerradas proporciona homogeneidad en los registros para poderse agrupar posteriormente. Este tipo de respuestas permiten valorar el grado en que aparece la conducta a través de los distintos niveles de frecuencia. Por otro lado, el uso de códigos numéricos permitirá analizar los datos con la ayuda de un programa informático.

¿Cómo es un buen cuestionario?

- Un buen cuestionario debe tener un aspecto lo más atractivo posible.
- Ser corto.
- Con una redacción clara.
- Sin preguntas negativas.
- Las preguntas deben estar bien ordenadas por temas.
- Dentro de un tema, tiene que ir de las ideas más generales a las más particulares.
- Las preguntas nunca deben orientar la respuesta; el que responde nunca debe sentirse juzgado.

Si desea ver cuestionarios enteros, puede consultar el utilizado por el Centro de Investigaciones Sociológicas para el barómetro de enero de 2001 activando el Caso 1.1 en la página web.

En la página web se dispone de un caso complementario (capítulo I, apartado 4.5).

A estas alturas de la película, se imaginará lo complejo que es hacer una encuesta que abarque todo el territorio nacional. Por ello, cuando hay posibilidades económicas, lo mejor es contratar una empresa especializada. La mayoría de los medios de comunicación recurren a esta estrategia. Esto fue también lo que hicieron los autores del estudio del Defensor del Pueblo.

Cuando no sea posible, los investigadores deberán formar adecuadamente a los entrevistadores sobre los aspectos específicos de la encuesta:

- el contenido de las preguntas,
- el tipo de respuestas,
- la forma de resolver las dudas,
- las características de los encuestados, y
- prever alguna forma de contacto inmediato para resolver imprevistos.

3.6. El trabajo de campo y el medio a partir del cual se encuesta

Aunque hasta ahora nos hemos referido a un caso donde la encuesta se realizó de forma personal –con un entrevistador presente–, hay otras formas de hacerlo: por correo, por teléfono (y por Internet).

Cuando hay un entrevistador que recoge las respuestas de los encuestados, puede aclarar alguna pregunta, asegurarse de que se contestan en orden, y animar a no dejar ninguna sin responder. Por otro lado, esto supone que hay que formar a los entrevistadores y asegurarse de que no se extralimitan en sus funciones orientando a los encuestados (*sesgo del entrevistador*). En conjunto, este medio es el más fiable aunque sea el más costoso.

Si se utiliza el correo, es todavía más importante que el cuestionario esté redactado de la forma más clara posible y que haya suficientes instrucciones, ejemplos, incluso algún teléfono de contacto para resolver dudas. Este medio es el más barato y evita la posible reactividad de los encuestados hacia el entrevistador. No resulta adecuado cuando deben representarse sectores de la población con escaso nivel cultural. Al no poder controlar el orden en que responden los encuestados, es posible que se produzca un *sesgo de respuesta*, al poderse conocer las últimas preguntas antes de proceder a contestar a las

primeras. La mayor dificultad de este medio estriba en la pérdida –no aleatoria– de encuestados. Es habitual que sólo responda un 50% de las personas, a pesar de enviarles un par de cartas adicionales a la encuesta, recordándoles lo importante que son sus respuestas y cuánto se lo agradecería la ciencia.

Las encuestas telefónicas combinan la "presencia" de un entrevistador y la economía de medios. Adicionalmente, el proceso de selección aleatoria, la marcación del número de teléfono y el registro y cómputo de datos pueden hacerse por ordenador, de manera integrada. Sin embargo, también tiene inconvenientes: la extensión del cuestionario tiene que ser bastante limitada, así como la profundidad de las preguntas. Puede existir mayor reactividad que con una entrevista personal, porque en esta última el encuestado escribe sin que sea testigo el encuestador; mientras que por teléfono uno debe responder a un desconocido. Por otro lado, este medio no es adecuado si deben ser representados sectores de la población que no tengan teléfono en su totalidad.

Las encuestas a usuarios de Internet, a través de la red, deben salvaguardar, como los otros medios, todas las exigencias de representación que hemos presentado. Una de las dificultades de esta modalidad es la falta de información sobre la población que va a ser encuestada; otra dificultad es la falta de datos sobre los posibles encuestados que no responden. Otra situación distinta es utilizar el correo electrónico como sustituto del correo con papel. No conocemos datos todavía sobre si el medio electrónico, por la rapidez y simplicidad, mejora las tasas de respuesta del tradicional; parece razonable pensar que es así y que su uso vaya en aumento.

Planificación de la investigación con encuestas. Alcance de las conclusiones.

Si accede al sitio del Centro de Investigaciones Sociológicas (<http://www.cis.es/>) podrá comprobar cómo, en sus encuestas periódicas, a veces se pregunta sobre un determinado tema de forma puntual y, otras veces, se pregunta sobre la misma cuestión en distintas encuestas. Eso refleja las dos estrategias generales en la investigación con encuestas: *diseño transversal* y *diseño longitudinal*.

En un *estudio transversal* se hace una fotografía, o una radiografía, de cómo es la realidad en un momento dado. Por ejemplo, "¿cuáles son los problemas que más preocupan a los españoles, en enero de 2006?", "¿qué puntuación le da usted hoy a tal líder político?", "¿ha dejado usted de consumir carne de pollo?". Son preguntas que retratan en las encuestas del CIS la realidad social en un momento dado.

Los estudios transversales no sólo describen las variables, también se pueden medir relaciones entre ellas, o establecer diferencias entre grupos.

Dentro de una encuesta pueden medirse relaciones, como, por ejemplo, la relación entre edad y hábitos de salud o la relación entre tabaquismo y grado de actividad física. Lo único que tiene que hacer el investigador es cruzar los datos de las preguntas que exploran cada una de estas informaciones.

Las tablas en las que se cruzan los resultados de dos preguntas se llaman de *contingencia*.

Para realizar una tabla de contingencia se prepara una cuadrícula donde las columnas corresponden a las posibles respuestas de una pregunta, y las filas a las posibles respuestas de la otra. Después, el investigador va situando a cada uno de los encuestados en la cuadrícula que le corresponde. Finalmente cuenta cuántos hay en cada una. En la Tabla 1.5 presentamos un ejemplo.

Tabla 1.5. Tabla de contingencia con dos preguntas para estudiar el grado de relación

		P.12. Sitúe usted el grado de actividad física en el que se incluye			
P.23. Indique el nivel de consumo de tabaco que le define		Nada	Ligero	Moderado	Mucho
	Nada	5	50	100	500
	Menos 1 paquete	10	100	50	25
	1 paquete	100	50	25	10
	>1 paquete	150	25	5	0

También pueden medirse diferencias entre grupos, por ejemplo, se puede contestar a la pregunta "¿leen más las mujeres que los hombres?" o "¿son más murmuradores los alumnos de centros privados que los de los públicos?". Fíjese en que estas relaciones o diferencias no surgen de forma espontánea del cuestionario ni son sugerencias del programa de análisis estadístico. Son subobjetivos cuidadosamente planificados por los investigadores.

En un *diseño longitudinal*, lo que más interesa estudiar es el posible cambio en las personas.

Las siguientes preguntas son ejemplos de diseños longitudinales: "¿ha variado la intención de voto?", "¿han cambiado las prácticas de seguridad en las relaciones sexuales?", "¿ha descendido el consumo de carne de pollo?". Para estudiar el cambio hay que describir a la misma población de la misma manera. No pueden introducirse "mejoras" en el cuestionario, porque, de hacerlo así, siempre planearía la amenaza de que un cambio en una variable puede deberse a la distinta forma de preguntar. De igual manera, no se deben introducir modificaciones en cuanto a los estratos representados o a las eta-

pas utilizadas en el muestreo. En resumen, si todo lo hacemos igual esperamos que las diferencias en las respuestas se deban a los cambios producidos en la población.

Sólo cuando se ha utilizado un tamaño muestral pequeño –siendo un tamaño adecuado– podemos plantearnos volver a preguntar a los mismos participantes (estudio de panel). Sin embargo, lo más habitual es realizar un nuevo muestreo. Si la población ha variado, ese cambio se detectará mediante una muestra que la represente correctamente.

En los estudios de panel se rastrea de forma más exhaustiva el cambio, ya que son exactamente los mismos encuestados los que responden dos veces.

Los estudios de panel se hacen más fácilmente con poblaciones pequeñas, donde se tienen localizados a los componentes de la muestra. Una universidad, una empresa, pueden ser contextos para un estudio de panel.

Existe otra variedad de estudio longitudinal, llamada *diseño de cohorte*, que consiste en comparar la forma en que cambian varias generaciones. Estos diseños de cohortes se llevan a cabo cuando alguna circunstancia temporal se liga a una generación (o promoción). Por ejemplo, puede tener sentido comparar la evolución en sus estudios de los primeros psicólogos que estudiaron siguiendo los planes de la ESO con una generación anterior de BUP. O comparar la evolución de la primera generación nacida bajo el euro con la evolución de la última generación de la peseta.

En los diseños de panel pueden estudiarse tanto diferencias *intra* cohorte a lo largo del tiempo, como diferencias *inter* cohorte en un momento dado.

A continuación presentamos ejemplos de estudios de panel: a) *inter cohorte*: diferencias en las notas (opiniones políticas, lecturas, aficiones...) en el primer año en las dos promociones; b) *intra cohorte*: diferencias de las notas (opiniones políticas, lecturas, aficiones ...) a lo largo de los años de la licenciatura.

En la página web se dispone de un caso complementario (capítulo I, apartado 4.6).

Si volvemos a observar la información de la Tabla 1.5 nos daremos cuenta de que no sabemos qué puede producir qué, ¿los que tenían el hábito de hacer ejercicio físico se sintieron menos inclinados a fumar? o ¿los que se iniciaron en el hábito de fumar no sentían ganas de hacer ejercicio?

En una encuesta no podemos saber la dirección de la acción de las variables.

También hemos de considerar que puede existir una tercera variable, no computada, que correlacione con las dos anteriores y que pueda explicar el hábito de fumar y hacer ejercicio; por ejemplo, haber tenido como modelo pa-

dres fumadores no deportistas o padres no fumadores deportistas. En los estudios descriptivos con encuestas, y en general en todos aquellos en los que los investigadores no tienen la capacidad de "producir" la acción de una variable sobre otra, sólo podemos concluir sobre datos descriptivos, diferencias entre grupos o relaciones entre variables.

En un estudio con encuestas no se puede concluir sobre la causa o influencia de una variable sobre otra.

No podemos acabar de estudiar este capítulo sin sugerirle que acceda al sitio del Instituto Nacional de Estadística (<http://www.ine.es/>), encargado de encuestas tan trascendentes como la que proporciona los datos para el IPC o la Encuesta sobre la Población Activa, que evalúa la situación laboral de los ciudadanos.

Capítulo II

Metodologías descriptivas II: la observación

1. La observación al servicio de la Psicología científica

1.1. Presentación de un caso

¿Viaja mucho en avión? Si no lo hace, le podemos asegurar que las terminales de los grandes aeropuertos internacionales son lugares en los que pueden observarse gran cantidad de curiosidades relativas al comportamiento humano. No vamos a hacer una lista de las que más nos llaman la atención. Simplemente le vamos a contar una historia que se desarrolla en la zona de llegadas. Allí se producen las conductas propias de personas que se encuentran después de haber viajado grandes distancias. Turnbull, Stein y Lucas (1995) consideraron que éste era el lugar idóneo para observar el modo en que la gente se abraza cuando tiene un encuentro emotivo.

Para que entienda las razones que tenían estos investigadores sudafricanos para elegir ese contexto de investigación necesitamos contar más cosas. Estaban interesados en estudiar el modo en que la especialización hemisférica puede llegar a explicar algunos de los comportamientos propios de las relaciones emocionales entre humanos. Por ejemplo, parece constatado que, tanto entre los humanos como entre los primates superiores, las madres –y los padres, cuando se animan a ello– sostienen sus bebés con el brazo izquierdo. Una primera razón aducida por los estudiosos del tema fue que eso permitiría liberar la mano derecha, cuya predominancia está más extendida. Sin embargo, se ha visto que no hay una correlación sistemática entre ser diestro y sostener al bebé con la izquierda y ser zurdo y sostenerlo con la derecha. Otra explicación más reciente aduce que la razón para que el bebé sea sostenido con el brazo izquierdo radica en que es el hemisferio cerebral derecho el encargado de procesar la información de tipo emocional. Así, la madre utilizaría sus receptores visuales y sensoriales del lado izquierdo (que son los que llegan más

rápidamente al hemisferio derecho) para la tarea de transmitir cuidado y cariño a su bebé y percibir sus expresiones y necesidades de forma más eficiente y rápida.

Turnbull, Stein y Lucas (1995) hicieron la conjetura siguiente: si es cierto que a los bebés se les sostiene con el brazo izquierdo debido a la especialización hemisférica, toda conducta que implique un procesamiento emocional y necesite una lateralización debería mostrar también una clara predominancia izquierda. ¿Se le ocurre en qué conducta pensaron que cumpliera estos requisitos? Exacto, los abrazos implican procesamiento emocional y suelen conllevar una lateralización. (¿Por qué lado cruza la cabeza con su pareja cuando la abraza? Si no lo recuerda ya tiene otra razón más para abrazarla en su próximo encuentro...)

Por eso se fueron al aeropuerto internacional más próximo a su universidad: para observar cómo se abraza la gente.

1.2. Observación científica: sistemática y replicable

Ahora bien, eso de ir por los aeropuertos espiando el modo de abrazarse de los viajeros no parece una tarea muy propia de investigadores científicos. Sueña más cercano al tipo de tareas que llevan a cabo los periodistas de los programas televisivos "del corazón". ¿Qué hace que la observación se convierta en científica? En nuestra opinión, dos son las claves: el contenido y la forma. Por un lado, el propósito de la observación tiene que ver con el incremento del conocimiento dentro de disciplinas científicas y no con la satisfacción de la curiosidad de los telespectadores. Por otro lado, la forma de llevar a cabo la observación en el contexto de la investigación científica conlleva unas garantías que no se dan habitualmente en otros contextos. Como ésta es una obra dedicada a las metodologías de investigación, aquí vamos a extendernos sobre la segunda clave.

Desde el punto de vista de la metodología de investigación, una *observación científica* debe cumplir las condiciones de ser sistemática y replicable. La observación es sistemática cuando se van haciendo explícitos cada uno de los pasos del proceso. Cuando ello es así, se está facilitando que también sea replicable, es decir, que pueda ser reproducida –y, por tanto, contrastada– de forma independiente en todos sus elementos.

En el apartado de "Método" del estudio llevado a cabo por Turnbull, Stein y Lucas (1995), se establece lo siguiente:

"Se observó a adultos abrazándose en la sala de llegadas de un aeropuerto internacional. Los participantes no sabían que se les estaba observando. Se definió *abrazo* como una interacción en la que un participante A, cuando se encuentra con B, lo abraza poniendo su cabeza en uno de los hombros de B y rodeándolo con los dos brazos. Se excluyeron los casos en los que se utilizaba sólo un brazo o en los que, aunque se cogieran por los hombros, no se daba el apoyo de la cabeza. Los abrazos a niños tampoco se tuvieron en cuenta [...] Se observó un total de 321 abrazos: 114 mujer/mujer, 174 mujer/varón y 33 varón/varón" (p. 18).

Este breve párrafo deja traslucir que las observaciones a que ha dado lugar el estudio que se presenta se han llevado a cabo utilizando un determinado procedimiento que empieza por delimitar el objeto de la observación –los abrazos–, el nivel de análisis de la observación –las díadas adulto/adulto–, los lugares de la observación –la sala de llegadas de un aeropuerto internacional–, etc. En la medida en que lo que se esboza se vaya concretando, cualquiera de nosotros, lectores de su trabajo, supuestas las habilidades y formación necesarias, estaríamos en condiciones de repetirlo, es decir, de replicarlo. Desmenuzar ese proceso de investigación es el objetivo de los apartados que siguen. A continuación detallamos los elementos del proceso para posteriormente ir desarrollando cada uno a lo largo de todo el capítulo.

1.3. Diferentes elementos del proceso de observación

Algunas palabras sacadas del lenguaje cotidiano sirven para hacerse una idea cabal acerca de los elementos del proceso de observación sobre los que es necesario que el investigador sea lo más sistemático posible: *qué* observar, *cómo* observar, *a quién*, *cuándo* y *dónde* hacerlo.

1) *Qué* observar.

Ya ha quedado claro que, en el ejemplo que nos ocupa, el objetivo es describir el modo en que los adultos se abrazan analizándolo desde la perspectiva de la lateralización del abrazo. Por tanto, "la lateralización del abrazo entre adultos" constituye el *qué* observar. Para ser sistemáticos respecto a su delimitación, es necesario tener en cuenta dos aspectos: el *marco teórico* y el *nivel de análisis*. Dada la importancia que tienen ambos aspectos, se les dedica el próximo apar-

tado. Baste con adelantar que la definición del objeto de observación conlleva una asunción teórica –el marco– y la delimitación de un plano o planos del fenómeno que se quiere observar –el nivel de análisis.

2) *Cómo* observar.

El elemento del proceso de observación relativo al *cómo* observar trata de delimitar la modalidad elegida para llevarlo a cabo, así como el tipo de registro que se va a utilizar. Ambos aspectos serán desarrollados con profusión en próximos apartados de este capítulo. Podemos anticipar que el primero hace referencia a cómo se sitúa el investigador frente el fenómeno que va a observar y el segundo al modo en que va a dejar constancia de su trabajo. Para ambos aspectos se abren diferentes posibilidades y todo investigador debe conocerlas para poder elegir entre ellas y hacerlas explícitas.

Veamos ahora cómo se delimitan el resto de los elementos señalados: *a quién, cuándo y dónde* observar. Seguimos con la investigación sobre "la lateralización del abrazo entre adultos".

3) *A quién* observar.

En el propio título que hemos escogido para referirnos a esta investigación está implícito que se eligió a personas adultas. En el párrafo que hemos transcrito literalmente se especifica algo más: no se consideraron los abrazos en que estaban implicados niños y se eligieron 321 parejas. Todo ello nos informa sobre *a quién* se observó. Note que no se nos relata de forma explícita el modo en que fueron elegidos. Por el relato del escenario parece innecesario. ¿Cree que los autores utilizaron algún procedimiento aleatorio de selección de la muestra al estilo de los presentados en el apartado de encuestas del capítulo anterior?

En la página web se dispone de una actividad complementaria (capítulo II, apartado 1.3).

Cuando lea cómo se elabora un *informe de investigación* en el Apéndice B, verá que las personas a las que se encuesta, a las que se observa o a las que se aplican las tareas experimentales son un elemento central en el proceso de investigación. Según los objetivos que se persigan y el objeto de la investigación, el modo en que se decide quién participa en la misma puede variar.

4) *Cuándo* observar.

Si repasa el párrafo en que se aludía al método seguido en la observación, notará que no hay ninguna alusión al momento en que se observó. A nosotros nos extraña esa omisión. Y vamos a explicar por qué. Si alguna vez ha hecho un largo viaje en avión, de esos que obligan a cambiar el horario y a pasar una noche dentro de la aeronave, habrá advertido que su capacidad de expresar

emociones en la terminal de llegada está seriamente mermada. Muy probablemente, en vez de abrazos, dé cabezadas... Haber especificado el tipo de vuelo y las horas de diferencia con los lugares de procedencia, en caso de haberse podido hacer, ayudaría a replicar mejor la investigación.

En general, cuando un investigador aborda la selección de los momentos para la observación, debe tener en cuenta si es relevante para su objetivo el momento en que ésta se lleva a cabo. Si es relevante, deberá especificar el modo mediante el cual ha decidido observar en unos momentos y no en otros.

Pero hay otra cuestión importante dentro del elemento relativo a *cuándo* observar. En la actualidad, excepto si resultan muy intrusivos, los sistemas de grabación audiovisual suelen ser el soporte en que se recoge –y se almacena– el producto de la observación. Pero aunque parezca que, de toda la vida, todo el mundo ha manejado el vídeo como maneja el grifo del lavabo, no hace mucho tiempo las observaciones se registraban *in vivo* en tiempo real. Es decir, el observador tenía que decidir sobre la marcha qué es lo que estaba pasando y dejar constancia de ello. A la hora de decidir si la observación era fiable podían surgir muchos problemas.

Por ejemplo, el observador A capta algo que califica como abrazo y pasa a registrarlo aunque el observador B no lo considere del mismo modo. Para ello necesita fijar su vista en el cuaderno de notas o la hoja de registro. Mientras lo hace, aparece una nueva pareja abrazándose que esta vez sí capta el observador B. El primer observador no puede hacerlo porque estaba ocupado en el registro del abrazo anterior. De ese modo, pueden encontrarse con que, habiendo estado observando a los viajeros que llegan durante los mismos minutos, sus datos parecen pertenecer a dos sesiones diferentes. Ese modo de observar registrando en cada momento lo que acontece se llama *registro continuo*. El problema que puede generar su uso se aborda haciendo un *registro por intervalos*. La sesión se divide en fracciones muy cortas de tal modo que se alternan los tiempos para mirar y los tiempos para registrar. Por ejemplo, se establecen intervalos de 2,5 minutos. Durante los dos primeros minutos se observa y el medio minuto restante se dedica a registrar lo observado. Todo esto es innecesario cuando se ha hecho una grabación en vídeo. A la hora de codificar puede volverse al punto de la grabación donde se empezó a registrar.

5) *Dónde* observar.

En el ejemplo que estamos utilizando no queda ninguna duda de que se observó en la sala de llegadas de un aeropuerto internacional y, como no conocemos sus dimensiones, pensamos que sólo había una. En toda investigación

en la que pueda observarse el fenómeno que se estudia en diferentes situaciones, el investigador deberá plantearse la selección de aquellas que más le interesen y presentar el modo de decidirlo.

Hasta aquí, la presentación de algunos de los elementos que configuran el proceso de observación cuando se cumplen los requisitos de ser sistemático y replicable.

Realizar un trabajo sistemático facilita que la observación sea replicable, haciendo explícitas sus opciones en relación con los diferentes elementos del proceso de observación. Si consigue ambas cosas, su observación podrá considerarse científica.

1.4. El objeto de observación: marco teórico y nivel de análisis

Al introducir nuestro ejemplo al principio del capítulo, explicamos cuál había sido el objetivo de la investigación: describir la lateralización en los abrazos entre adultos como modo de generalizar la hipótesis de la dominancia hemisférica derecha en el procesamiento de los indicadores emocionales. Podríamos decir que se está abordando el problema del comportamiento en contextos emocionales dentro de un marco teórico sociobiológico. De forma muy breve, eso quiere decir que los autores se sitúan en una tradición de investigación en Psicología y Ciencias Sociales en la que se utilizan como principios explicativos elementos teóricos tomados de la Biología evolucionista. La conducta está regulada biológicamente mediante procesos que han ido originándose por la selección natural.

Si es un entusiasta de las ciencias de la vida y de las explicaciones evolucionistas, este tipo de marco teórico le parecerá muy adecuado. Pero estará de acuerdo con nosotros cuando afirmamos que no es el único desde el cual puede investigarse la conducta del abrazo. Si cree que toda conducta puede ser explicada mediante la reconstrucción del proceso de aprendizaje que la originó, se situaría dentro de otro marco teórico: el conductismo.

No queremos ahorrarle el trabajo de cursar las asignaturas básicas de contenidos en Psicología, lo que queremos es que caiga en la cuenta de que un sociobiólogo y un conductista mirando a los viajeros que se abrazan en la sala de llegadas de un aeropuerto, aun mirando lo mismo, observarán cosas diferentes, ya que su marco teórico condiciona –orienta– su proceso de observación.

Por cierto, con los resultados no se pudo apoyar la hipótesis de la asimetría en el abrazo como extensión de la asimetría del procesamiento emocional a la hora de sostener a los bebés en brazos. Ninguna de las combinaciones de días entre varones y mujeres mostró preferencia por la izquierda. Sólo en el caso de los abrazos entre mujeres se dio la preferencia contraria: cruzaban sus cabezas por la derecha. Esto lleva a los autores a concluir que "el abrazo orientado a la derecha puede ser producto de alguna preferencia cultural específica" (Turnbull, Stein y Lucas, 1995, p. 20). A nosotros se nos ocurre que la extensión de la situación del cuidado del bebé al abrazo entre adultos era bastante arriesgada y, además, resultaba equívoca para predecir el comportamiento de adultos zurdos, ya que, por ser minoría, aprenden a abrazarse por donde lo hace la mayoría diestra.

Además de conocer el marco desde el que vamos a observar, es necesario dar un paso más: establecer el plano, delimitar el *zoom*. Eso es el *nivel de análisis*. En el ejemplo de los abrazos en los aeropuertos, el nivel de análisis está circunscrito a la díada adulto-adulto, con las tres combinaciones que se dan al tener en cuenta el género de las personas que se abrazan. En un laboratorio se podría, además, haber registrado la actividad cerebral de los dos hemisferios para ver si, además de la lateralización en el cruce de cabezas, se da una activación diferencial, tal como sostienen los autores del trabajo. Ese cambio de la observación de la interacción a la observación de la actividad cerebral podría asemejarse al *zoom* de una cámara que permite acercar o alejar el plano según los propósitos de la grabación.

Aunque la metáfora del *zoom* puede haberle llevado a pensar que los niveles de análisis, como los planos de una filmación, pueden ser infinitos, en Psicología y otras disciplinas afines nos movemos dentro de un número limitado de ellos. A continuación encontrará dos propuestas diferentes de posibles niveles de análisis tomadas de León y Montero (2003) y de Rosa (2000).

Diferentes propuestas de niveles de análisis en Psicología

Aunque elaboradas en contextos diferentes, las dos propuestas que presentamos sirven para ilustrar las posibilidades en que nos movemos dentro del ámbito de la Psicología. Ambas deben entenderse como orientativas, nunca como definitivas.

Rosa (2000) establece cuatro niveles:

- 1) **Nivel ecológico-social:** se incluye en este nivel el análisis de las acciones externas realizadas en un entorno físico y social.
- 2) **Nivel individual:** acciones internas.
- 3) **Nivel computacional:** activaciones de redes neuronales que se interpretan mediante modelos computacionales.
- 4) **Nivel biológico:** funciones llevadas a cabo por estructuras biológicas.

La segunda propuesta, realizada por León y Montero (2003), establece siete niveles:

- 1) **Especie:** humana, primates superiores, animales de laboratorio, especies en cautiverio, etc.
- 2) **Cultura:** anglosajona, latina, magrebí, subsahariana, etc.
- 3) **Grupo:** familia, pandilla, equipo de trabajo, comunidad religiosa, etc.
- 4) **Díada:** pareja, madre/hijo, profesor/alumno, etc.
- 5) **Individuo:** adultos, niños, estudiantes universitarios, ancianos, etc.
- 6) **Proceso psicológico:** personalidad, atención, memoria, emoción, motivación, etc.
- 7) **Componente psicológico:** conducta manifiesta, computaciones, activaciones fisiológicas, redes neuronales, inconsciente, etc.

Es obligación para todo investigador referirse explícitamente al marco teórico en que se mueve y a los niveles de análisis que elige para que sus investigaciones puedan ser replicadas.

2. Diferentes modos de plantear la observación

2.1. Tipos de observación

En el apartado anterior, cuando se describían los elementos que configuran el proceso de observación, se ha aludido al contenido de este epígrafe. Decíamos que el investigador tiene que elegir entre varias opciones el modo en que se sitúa –mentalmente– frente a lo que va a observar. Aunque diferentes autores abordan la clasificación de estas opciones de diferente modo (León y Montero, 2003; Shaughnessy, Zechmeister y Zechmeister, 2003), dichas opciones están claramente delimitadas. Nosotros vamos a presentar modos diferentes de llevar a cabo la observación. Después de describir las características de cada uno de ellos, le presentaremos los criterios para su posible clasificación.

Recuerde cómo se planteó la investigación del estudio descriptivo "la lateralización de los abrazos". Se trataba de observar a las parejas de adultos en la zona de llegadas de un aeropuerto. Hasta aquí, todo estaba claro. Pregúntese ahora en qué grado el investigador creaba el contexto de la observación. Parece que, en nuestro caso, el investigador no aportaba nada más que su presencia al contexto: el trajín de personas en la sala de llegadas se produce sin que

el investigador intervenga en modo alguno. Él se limita a registrar aquello que le interesa dentro de un conjunto de cosas que ocurren sin que él intervenga de ninguna forma.

Cuando esto es así, cuando el investigador se limita a observar una situación sin aportar nada a su creación –más allá de su ineludible presencia–, decimos que está haciendo *observación natural*.

La observación natural tiene la ventaja de permitir el acceso a las situaciones tal y como se producen habitualmente. Pero muchas veces el investigador no puede estar esperando a que, espontáneamente, aparezca el fenómeno que le interesa. Piense en el fenómeno del habla privada infantil, también conocida como *habla egocéntrica*. Se ha constatado que, entre los tres y seis años, los niños hablan para sí mismos con mucha más frecuencia de lo que lo harán según vayan creciendo. El origen y funciones de ese hablar para sí ha dado lugar a mucha investigación ¹¹.

Por contra, uno de los mayores riesgos para la investigación de este fenómeno es el de acceder a él: el mero hecho de ponerse a observar lo que hacen niños de esas edades no garantiza que manifiesten habla privada. El que ésta se ponga en funcionamiento depende de muchas cosas: la demanda de la tarea que se realiza, su contenido, el nivel de dificultad, el contexto, la cercanía de otras personas, el modo de interacción con las mismas... y así hasta un largo etcétera. Pues bien, un modo que tenemos los investigadores para, manteniendo el contexto natural de aparición del fenómeno, aumentar la probabilidad de que éste aparezca consiste en introducir algunos elementos en la situación, crear un contexto estructurado que así lo permita. Por ejemplo, introduciendo un juego que sepamos que resulta motivador e implica un reto para los niños que van a ser observados. En esas condiciones (Díaz y Berk, 1992), se sabe que aumenta la probabilidad de que se emita lenguaje autodirigido.

Cuando eso ocurre hay un cierto grado de intervención en la situación, lo que se observa está de algún modo provocado por el investigador. Esta modalidad se denomina *observación estructurada*.

Pero todavía hay más. Una cosa es provocar un fenómeno para describirlo y otra cosa es provocarlo para saber qué lo causa. Cuándo en el contexto de la metodología de investigación hablamos de *experimento* nos referimos al contraste de una relación de causa-efecto entre dos variables. En el capítulo III, tendrá ocasión de estudiar todo lo relativo a este asunto. Cuando la observación se estructura para llevar a cabo un proceso de contraste de relaciones causales, hablamos de *experimento de campo*.

11. Puede consultar Díaz y Berk (1992).

El criterio que nos ha permitido distinguir entre las tres modalidades de observación hasta ahora presentadas ha sido el del grado de intervención del investigador en la situación observada. Pero hay otro criterio que permite distinguir entre formas de observación: llevarla a cabo desde fuera o llevarla a cabo desde dentro. Es decir, el observador puede ser alguien ajeno al objeto de la observación o puede ser parte de lo que se observa. Cuando se elige esta segunda opción hablamos de *observación participante*. Cuando no es así se habla de *observación externa*.

En la página web se dispone de una actividad complementaria (capítulo II, apartado 2.1).

Se habrá dado cuenta de que, en realidad, le hemos presentado dos sistemas de clasificación de los posibles modos de enfocar la observación. El primero de ellos hace referencia al grado de estructuración de la situación que se observa. Da lugar a tres modalidades: *observación natural*, *observación estructurada* y *experimento de campo*. El segundo es relativo al grado de participación del observador en el fenómeno observado. Da lugar a dos posibilidades: observación externa y observación participante. En la tabla 2.1 le presentamos las seis combinaciones resultantes.

Tabla 2.1. Diferentes modalidades de observación

		Grado de estructuración		
		Natural	Estructurada	Experimento
Grado de participación	Externa	Externa natural	Externa estructurada	Externa experimento
	Participante	Participante natural	Participante estructurada	Participante experimento

Lo que queremos recalcar ahora es que cualquier estudio hecho mediante observación puede ser clasificado *a la vez* según los dos criterios. Por ejemplo, una observación natural puede ser participante o mediante observador externo. O visto de otro modo, el hecho de que sepamos que una observación es estructurada no nos dice nada respecto a si el observador era externo o participante.

En la actividad 2.1 de la página web tendrá ocasión de practicar acerca de los sistemas de clasificación observando fragmentos de grabaciones pertenecientes a cuatro estudios diferentes. Los dos primeros realizados por otros colegas: Español (2000) y Español y Rivière (2000). Los otros dos de diseño propio: Montero, Pérez del Palacio y González (2001).

En la página web se dispone de una actividad complementaria (capítulo II, apartado 2.1).

2.2. Tipo de código

Aunque desde hace unos años se utilicen los soportes audiovisuales para dejar constancia de las observaciones realizadas en un estudio, parece claro que un montón de cintas de vídeo, o de discos compactos, no constituyen los hallazgos de una investigación descriptiva llevada a cabo mediante observación. Una investigación se inicia con unos objetivos, dentro de un marco teórico, y delimita un nivel de análisis o varios. De todo lo que puede apreciarse en las grabaciones que ha presenciado hasta ahora –si ha hecho las actividades propuestas–, los investigadores sólo se han preocupado de algunas cosas: sus categorías de observación. Éstas se ven reflejadas en los registros que han llevado a cabo. Dichos registros se realizan utilizando códigos previamente establecidos. Cabe apuntar, por tanto, que en principio puede haber tantos códigos como investigaciones. Sin embargo, aquí nos interesa resaltar que todos ellos pueden reducirse a dos tipos: *registro narrativo* o *código arbitrario*.

Decimos que un observador utiliza un registro narrativo cuando se limita a utilizar el lenguaje natural como vehículo mediante el cual expresar sus categorías de análisis y dejar constancia de su aparición en los diferentes momentos de la observación.

En la actividad 2.1 de la página web hemos omitido el tipo de código, pero podrá acceder a él en la próxima que le vamos a proponer en breve.

Aunque en sentido estricto el lenguaje natural también lo es, hablamos de *código arbitrario* cuando el observador crea para la ocasión el conjunto de categorías que va a utilizar para dejar constancia de lo acaecido.

La definición anterior no implica que cada investigador "invente" su propio código. Dentro de una misma área de investigación, se suele compartir un conjunto de categorías para dar cuenta de los fenómenos de interés dentro de la misma. Ése es el caso de la investigación del fenómeno del habla privada dentro del aula.

En la actividad 2.2 de la página web le enseñamos a utilizar una versión simplificada del código que se utiliza en ese campo. Recuerde que también podrá ver un caso de registro narrativo y notar la diferencia entre ambos tipos de código.

En la página web se dispone de una actividad complementaria (capítulo II, apartado 2.2).

3. Construcción de códigos

3.1. Pasos en la elaboración de un código

Cuando hemos hablado de tipos de registro, al referirnos al uso de *códigos arbitrarios*, se ha dado a entender que dentro de cada ámbito de investigación se utilizan conjuntos de categorías ya establecidos. En nuestra propia experiencia como investigadores, eso fue lo que encontramos cuando empezamos a investigar sobre el fenómeno del habla privada. Como nosotros, la mayoría de los investigadores utilizan un código ya depurado por Berk (1986).

Pero alguien tiene que ser el primero. Cuando un nuevo fenómeno se empieza a investigar, las categorías todavía están por elaborar. En este apartado queremos mostrar lo cuidadoso que se debe ser cuando tiene que afrontarse la tarea de elaborar un nuevo código de observación. Para ello, vamos a apoyarnos en las recomendaciones que se ofrecen en un libro que puede considerarse un clásico en esta especialidad metodológica. Nos referimos a la obra de Bakeman y Gottman (1986). Estos autores recomiendan seguir los consejos siguientes:

- 1) Tener, al menos, una pregunta antes de empezar.
- 2) Elegir uno o varios niveles de análisis para comenzar a observar.
- 3) Hacer una observación previa asistemática.
- 4) Elegir categorías con suficiente nivel de detalle.
- 5) Elegirlas de modo que sean exhaustivas y excluyentes.
- 6) Someterlas a depuración.

Vamos a dedicar un breve espacio a cada uno de estos consejos. Los hemos agrupado en dos grandes conjuntos porque nos parece que las tres primeras recomendaciones pueden configurar un primer paso del proceso, mientras que las otras tres se deberán llevar a cabo en un momento posterior.

3.2. La pregunta, el nivel de análisis y la observación previa

Para hacer su tarea de aprender más relajada, pero no por ello menos eficaz, le invitamos a seguir nuestra explicación dentro de un contexto que no siendo psicológico puede resultar lo suficientemente familiar y cercano.

¿Conoce la palabra *asistencia*? El diccionario de la Real Academia de la Lengua recoge once acepciones de la palabra, pero ninguna de ellas es la que nos interesa. En el diccionario de uso de María Moliner, la palabra aparece también con once acepciones. La novena es la que nos interesa: "En baloncesto, pase que permite conseguir una canasta". Esperemos que no le desagrade este juego deportivo porque vamos a convertirlo en nuestro cómplice didáctico durante un espacio importante de nuestra lección. Lo que vamos a hacer es presentar el código a que pertenece la categoría "asistencia" de tal modo que entienda que quienes lo diseñaron siguieron los seis consejos reseñados más arriba. Pero empecemos por el principio.

3.2.1. La pregunta

El consejo recogido en primer lugar hace referencia a la necesidad de tener, al menos, una pregunta que responder para empezar a observar algo.

Para quien presencia un partido de baloncesto, una pregunta relevante puede ser quién es el mejor jugador del partido. De hecho nos atrevemos a aventurar que eso se preguntan los ojeadores de los equipos de la liga profesional americana cuando hacen su trabajo para, al final de la temporada, ofrecer informes a los equipos sobre los jugadores que terminan la liga universitaria para contratarlos en la profesional, después del famoso proceso de *drafts* (rondas de selección). En ese contexto es donde se han elaborado el conjunto de categorías que se usan para establecer la calidad de un jugador, lo que se conoce como sus estadísticas personales. Una curiosidad de ese proceso es que

el equipo que primero puede elegir es el que peores resultados obtuvo en la temporada anterior. Eso facilita la máxima igualdad entre los equipos y una incertidumbre y emoción más grande en cada liga.

Probablemente el juego le parezca tan conocido que no puede imaginar a nadie que, presenciando un partido por primera vez, no capte la importancia de la categoría "asistencia". Seguro que lo imagina mejor si se sitúa en el contexto de un partido de béisbol: ¿cómo hacerse idea de la importancia de los *strikes* para el equipo que defiende?, ¿pero qué equipo defiende y qué equipo ataca? ¿No cree que una persona puede dedicar muchas horas a ver partidos de ese deporte y seguir sin ser capaz de entender –observar– nada que tenga sentido? Si no se conoce el objetivo del juego –si no se tiene una pregunta– no pueden elaborarse un conjunto de categorías relevantes para su análisis –para su respuesta.

Imaginar a un aficionado novato aprendiendo las reglas del juego a base de ver partidos nos sirve como alegoría para ilustrar lo que ocurre cuando se está empezando a describir un nuevo fenómeno mediante observación. Ni el aficionado ni el investigador verán nada por mucho que miren. Necesitan una pregunta previa: ¿quién es el mejor jugador?

3.2.2. El nivel de análisis

Recordará que para explicar el concepto de *nivel de análisis*, introdujimos la idea de diferentes planos para observar el mismo fenómeno.

Pues bien, la conexión con el baloncesto es fácil si se sitúa en el papel del realizador de la transmisión de un partido en el que se va a decidir quién es el mejor jugador. La pregunta sería: ¿con qué plano, o planos, se capta la calidad del jugador? O más concretamente, ¿qué plano, o planos, necesitamos mantener para poder observar a un jugador? Piense que los realizadores no tienen infinitas cámaras, aunque sea posible pensar en un número ilimitado de planos. Es fácil deducir que los primeros planos del rostro de un jugador no serán necesarios para ilustrar su calidad –lo cual no quiere decir que no sirvan para ilustrar otros aspectos tales como el grado de concentración, el dolor tras un golpe, etc.–. Lo mismo cabe decir de planos aéreos. Pueden venir bien para ilustrar el ambiente entre el público que llena el Palau, pero no la calidad de los jugadores azulgranas. Parece que lo mejor es dedicar una cámara a seguir los pasos del jugador elegido y de sus compañeros manteniendo planos que permitan captar las incidencias: pases, puntos, faltas, etc.

En cualquier caso queda claro que con un único plano no tiene por qué agotarse el modo de ilustrar la calidad de un jugador –si hablamos en general– o el número de asistencias –si nos referimos a un indicador más concreto–. Pueden –casi deberíamos decir, deben– utilizarse varios planos para un mismo proceso observacional. Sustituya el contenido deportivo del ejemplo por una dimensión de la conducta humana –la agresividad infantil, por ejemplo– y entenderá que este consejo acerca del número de planos es relevante para la descripción de cualquier fenómeno accesible a la observación. Para la traducción del concepto de plano filmográfico al de nivel de análisis en Psicología le vendrá bien repasar las diferentes propuestas de niveles de análisis en Psicología que se muestran en el apartado 1.4.

3.2.3. La observación previa asistemática

Piense en el primer ojeador de la historia del baloncesto. Sabe lo que está buscando, el mejor jugador. Sabe que tiene que fijarse en los que destaquen en su contribución al resultado, tanto en un plano estrictamente individual como en su aportación al rendimiento del resto de sus compañeros. Y se pregunta: ¿cuáles son los mejores indicadores de la calidad de un jugador?, ¿qué hace que un jugador destaque en un partido?

Algunos indicadores –como los puntos anotados– debieron aparecer enseguida en la cabeza de este imaginario personaje. Pero otros –quizá el número de asistencias fuera uno de ellos– debieron llevar un poco más de tiempo para emerger ante su mente. Hasta que empezó a utilizar las categorías definitivas, necesitó llevar a cabo observaciones de forma no sistemática. Mientras no tuvo el código arbitrario terminado, debió utilizar el lenguaje natural –el *registro narrativo*– para expresar lo que había observado. Algo así como un informe después del partido en el que especificara, por ejemplo:

"El jugador Fulanito de Tal destaca claramente en conseguir puntos desde debajo de la canasta, pero es muy poco efectivo cuando se le hace una defensa zonal, ya que su tiro de media distancia no es bueno. Cuando ataca, pelea mucho por los balones y atrapa muchos rebotes, pero es lento en la defensa y pelea menos bajo su aro".

Si nos fijamos, en el relato emergen categorías que pueden empezar a ser útiles y candidatas para formar parte del código definitivo: tiro de cerca, tiro de lejos, rebote ofensivo, rebote defensivo... Teniendo ya una primera versión del código –por ejemplo, la que aparece en la tabla 2.2– nuestro observador empezaría a probarlo de forma sistemática.

Tabla 2.2. Primera versión para un código de observación de la calidad de un baloncestista

Puntos de campo	
-----de tiro libre	
-----de tres	
Rebotes	
-----defensivos	
-----en ataque	
Asistencias	
Tiempo jugado	
Faltas	
-----personales	
-----técnicas	
Balones perdidos	
Puntos conseguidos por el jugador contrario al que se defiende	

De igual modo se hace en un proceso de investigación mediante observación. Para que se empiece a consolidar un primer conjunto de categorías como relevantes para el análisis de un nuevo fenómeno, se necesita un periodo de observación que necesariamente será previo a la construcción del código. Asimismo, al no tener otro modo de registro, el lenguaje natural es la forma de recoger lo acontecido.

Siguiendo estos tres consejos, cerramos un primer momento del proceso. Ya hemos delimitado el objetivo y los niveles de análisis. Tenemos también una primera propuesta de código: ¿cómo saber si es adecuada para una observación sistemática y replicable?

3.3. El primer conjunto de categorías exhaustivas, excluyentes y con suficiente nivel de detalle. Su depuración

Lo que hacemos ahora es investigar si el conjunto que hemos pensado como primera propuesta cumple una serie de condiciones, condiciones que se hacen imprescindibles para que el código pueda ser útil para fines científicos. Todas ellas se describen en las tres recomendaciones siguientes.

- Elegir categorías con suficiente nivel de detalle
- Elegir categorías de modo que sean exhaustivas y excluyentes
- Someterlas a depuración

1) Elegir categorías con suficiente nivel de detalle

Aunque hasta ahora, al seguir el ejemplo del baloncesto, el uso de los medios audiovisuales haya parecido obvio, aquí es necesario hacer la precisión de que esta recomendación es especialmente importante en el caso de hacer observación *in vivo*. Imagínese que el primer ojeador de buenos jugadores empezó su labor cuando todavía no se televisaban los partidos. Entonces no tenía registro audiovisual sobre el que resolver dudas o contrastar datos. Resulta que ha encontrado dos jugadores que han hecho el mismo número de anotaciones y ha registrado los puntos de cada uno de ellos. Cuando lleva su información al equipo técnico que decidirá sobre la contratación, surge la duda de cómo decidirse por uno de ellos, ya que el criterio fundamental es la eficacia anotadora. El entrenador no lo duda, propone elegir a aquel que haya anotado más tiros exteriores, ya que es más necesario para la actual composición del equipo. El ojeador se lleva las manos a la cabeza: ¡no ha registrado qué tipo de punto era el que se hacía cada vez! Al no establecer el suficiente nivel de detalle para el objetivo que tenía se ve abocado a una situación muy poco deseable. Deberá repetir todo el proceso. Aun en el caso de haber utilizado medios audiovisuales de grabación, estará obligado a volver a ver a todo el partido. ¿Qué hubiera pasado en caso contrario, si el ojeador recoge el detalle del tipo de tiro y el entrenador hace caso omiso de él? Nada. A lo sumo, el ojeador habría sentido que había trabajado más de la cuenta. Pero este inconveniente no tiene parangón con el anterior.

En la observación en otros contextos, más propios de la investigación académica, puede pasar exactamente igual, que necesitamos una información precisa que no hemos detallado suficientemente. ¿Hay manera de preverlo? No. Recuerde que estamos en un proceso de elaboración de un código para observar algo que se describe por primera vez.

La recomendación es que, en lo que se refiere al nivel de detalle, más vale pasarse que quedarse corto.

2) Elegir categorías de modo que sean exhaustivas y excluyentes

Además de cumplir la condición anterior, el grupo de categorías que se han derivado de la fase anterior deben ser exhaustivas y excluyentes.

En la página web se dispone de una actividad complementaria (capítulo II, apartado 3.3).

Que sean exhaustivas quiere decir que, en su conjunto, agoten todos los elementos del fenómeno en observación.

La otra condición, la de ser excluyentes, hace referencia a que un elemento del fenómeno en observación sólo pueda pertenecer a una de las categorías.

Aunque pueda parecerle que estas dos condiciones son fáciles de conseguir, hay muchos casos en los que no se cumplen. Cuando es así, los análisis de los datos recogidos en la investigación se dificultan mucho. No es el momento de profundizar más en el asunto. La moraleja simplemente va en la línea de enfatizar que cualquier conjunto de categorías que se nos ocurra no cumple *per se* estas dos condiciones.

3) Someterlas a depuración

Las condiciones que deben cumplir las categorías no se agotan con las mencionadas en los epígrafes anteriores. Establecer su cumplimiento parece una labor de reflexión, de despacho. Podemos decidir pensando que el conjunto de la Tabla 2.2 necesita añadir, al menos, la categoría "robo de balón", para que pueda considerarse exhaustivo. Del mismo modo, para que todas las categorías sean excluyentes entre sí, se hace necesario aclarar los límites entre "falta personal" y "pérdida de balón". Por ejemplo, "falta personal en ataque" –sin computar como pérdida de balón– y "falta personal en defensa".

Depurar el código representa preguntarse si cumple algunos requisitos que sólo pueden ser establecidos de forma empírica.

En el último apartado del capítulo se aborda la depuración del código. Para activar su curiosidad, pregúntese si considera igual de fácil que dos espectadores de un mismo partido se pongan de acuerdo en que un tiro ha sido de tres puntos o en que un punto haya sido obtenido mediante tiro libre. Por lo tanto, ¿cuál de las dos categorías –tiro de tres o tiro libre– cree que está más depurada?

4. Tipos de medida en la observación

4.1. Medidas: ocurrencia, frecuencia, latencia, intensidad, duración

Después de tanto baloncesto quizá tenga cierta sensación de cansancio físico. Si es así, no crea que el simple hecho de pensar en actividad física le produce fatiga. Si verdaderamente siente cansancio, se deberá, muy probablemente, al hecho de mantener tanto rato la misma postura en su lugar de estudio. Posiblemente le surjan molestias en las vértebras cervicales, las cuales acaban concentrándose en cierta sensación de dolor entre los dos omóplatos... Además, es muy probable que su circulación sufra dificultades en la zona abdominal y en las extremidades inferiores...

Piense: ¿tiene ese dolor cervical ahora? Veamos, durante la última semana, ¿cuántos días estuvo trabajando durante un periodo superior a media hora?, ¿en cuántos de ellos sintió molestias cervicales? Probablemente, su respuesta sea: "no me he *fijado*". Si se hubiera fijado, si lo hubiera *observado* y *registrado*, ahora podría contestar a nuestras preguntas sin ninguna dificultad.

Con este ejemplo, estamos poniendo de manifiesto que el proceso de observación necesita un último paso: la medida. Hay diversas maneras de dejar constancia de un fenómeno de interés para el observador. Le vamos a presentar cinco modos diferentes de medida en la observación:

- 1) Ocurrencia
- 2) Frecuencia
- 3) Latencia
- 4) Duración
- 5) Intensidad

La primera de las medidas, la ocurrencia, nos dice simplemente si una categoría ha aparecido durante un periodo de registro. Si retomamos la categoría de observación "tener dolor entre los omóplatos", la pregunta relativa a si lo padece ahora es una pregunta sobre su ocurrencia. Fíjese en que simplemente anotamos si algo ha ocurrido en un intervalo determinado. Sólo hay dos posibilidades de respuesta: sí, ha ocurrido (sí, tengo dolor entre los omóplatos); no, no ha ocurrido (no tengo dolor entre los omóplatos).

Imaginemos que es de las personas que responden que sí a la pregunta sobre el dolor postural. Desde luego, no es agradable sentir esa molestia, pero también está claro que no es lo mismo haberla padecido alguna vez que padecerla a menudo.

La ocurrencia nos informa sobre la aparición de la categoría que observamos, la frecuencia nos dice cuántas veces lo ha hecho. Al incluir la frecuencia como medida de lo que estamos observando, estamos recogiendo más información.

La frecuencia puede informarse de dos maneras: absoluta y relativa. Imagine que hay dos estudiantes que se quejan de haber padecido el dolor cervical dos veces en las últimas cuatro semanas. Aparentemente, la frecuencia con que padecen el dolor es la misma. Pero imagine que una de estas dos personas es muy aplicada y todos los días dedica más de dos horas al estudio, mientras que la otra, por las razones que sean, no puede estudiar más que una vez cada dos semanas. ¿Sigue pensando que las dos personas son igual de propensas a padecer dolor postural? Está claro que esas dos veces no significan lo mismo. En un caso implican tener dolor en una de cada catorce ocasiones mientras que en el otro significa tener el dolor cada vez que es posible que aparezca.

Cuando informábamos que dos personas habían respondido tener el dolor dos veces –sin ninguna referencia a las oportunidades de padecerlo– estábamos haciendo referencia a la frecuencia absoluta de aparición de la categoría bajo observación. Cuando hemos informado también de las oportunidades (dos de dos o dos de veintiocho), hemos hecho referencia a su frecuencia relativa.

Todavía podemos obtener información más rica sobre el problema de los dolores de espalda de las personas que estudian sentadas durante largos periodos de tiempo. Supongamos ahora que encontramos dos personas que padecen el dolor con la misma frecuencia relativa, uno de cada tres días de estudio sienten esos pinchazos entre los omóplatos. Sin embargo, la primera de ellas nos dice: "Es sentarme y... ¡empieza a dolerme!". La otra, por el contrario, afirma: "a mí me empieza a doler hacia el final, cuando llevo ya mucho rato trabajando". Ambas informan de la misma frecuencia pero de diferente latencia.

La *latencia* es el tiempo que tarda en aparecer la conducta que se observa a partir de un determinado punto de referencia.

En este ejemplo, es el momento de empezar a estudiar sentado ante la mesa de trabajo. El informe verbal que hemos recogido ha sido impreciso. Una buena investigación informaría de la latencia en unidades de tiempo. En algunos experimentos de psicología se mide en milésimas de segundo. Ése es el caso de los tiempos de reacción que son el tiempo que tarda un sujeto en responder a la presencia de un estímulo perceptivo.

Como habrá podido notar, cada una de las medidas que hemos ido utilizando para informar sobre la categoría que estamos observando nos ha permitido ir apreciando mejor la naturaleza del fenómeno y, sobre todo, la diferente gravedad del mismo. La siguiente medida casi no tiene ni que explicarse: ¿cuánto tiempo dura el dolor? Como habrá fácilmente adivinado, estamos hablando de duración. Independientemente de cuánto tiempo tarde en aparecer, no cabe duda de que un buen procedimiento para apreciar la gravedad del problema de los dolores de espalda es conocer su duración una vez que han aparecido.

Hablamos de duración para referirnos al tiempo en que una conducta bajo observación está presente de forma continuada.

Si alguna vez ha tenido la desgracia de padecer un dolor de espalda durante un periodo considerable de tiempo, se habrá dado cuenta de que, en algunas ocasiones, lo ha podido sobrellevar, mientras que, por el contrario, en otras ha tenido que parar e, incluso, tomar alguna medicación. Eso ha ocurrido cuando el dolor ha sido muy intenso.

Con la intensidad informamos sobre *el grado* en que la conducta bajo observación está presente. Para poder hacerlo necesitamos construir escalas de apreciación de esa intensidad.

Aunque le parezca que es fácil apreciar diferentes grados de intensidad del dolor de espalda, los psicólogos y otros científicos sociales llevamos muchos años dedicando muchos esfuerzos al arte de medir la intensidad de variables que son de apreciación subjetiva.

En la página web presentamos consejos para evitar los dolores de espalda.

En la página web se dispone de una animación complementaria (capítulo II, apartado 4.1).

Haciendo un rápido repaso a las medidas en la observación, recordemos que:

- La ocurrencia nos informa sobre la aparición de determinada categoría en el periodo de observación.
- La frecuencia nos indica, además, cuántas veces ha aparecido. Dicha frecuencia puede ser absoluta o relativa dependiendo de que la cantidad de veces en que aparece la categoría sea puesta, o no, en relación con la cantidad de veces que podría haber aparecido.
- La latencia nos señala el tiempo que transcurre desde un momento concreto –por ejemplo, la aparición de un estímulo o el comienzo de un intervalo– hasta la aparición de la conducta en cuestión.

- Hablamos de duración para referirnos al tiempo en que la categoría está presente de forma ininterrumpida.
- Por último, la intensidad nos informa del grado en que la conducta aparece.

En la página web se dispone de una actividad complementaria (capítulo II, apartado 4.1).

Habrás notado que en el deporte de las canastas las medidas a las que da lugar el uso de las categorías que permiten observar el rendimiento individual de un jugador suelen ser del tipo frecuencia. Tenga en cuenta, de todos modos, que no hemos puesto absolutamente todas las categorías que pueden utilizarse. Algunas, como las relativas al palmarés –por ejemplo, Epi fue medalla de plata en los Juegos Olímpicos de Los Angeles de 1984–, serían expresiones de ocurrencia. Asimismo, las puntuaciones conseguidas en un concurso de mates, donde unos jueces otorgan una calificación a la calidad del mate realizado por el concursante, serían medidas de intensidad. Algunas de las categorías menos conocidas son medidas de latencia, como el tiempo transcurrido hasta cometer la primera falta personal o hasta anotar la primera canasta.

Otro hecho que no queremos que le pase inadvertido es el de que no toda categoría puede ser expresada en todos los tipos de medida. El ejemplo del dolor de espalda es un caso en el que sí puede hacerse, pero las diferentes categorías sobre el rendimiento individual en baloncesto ilustran el caso contrario. Un tiro de tres puede ser expresado como ocurrencia, como frecuencia –absoluta y relativa– y como latencia, pero no como duración o intensidad.

Es decir, en algunos casos la elección del tipo de medida puede ser determinado por el investigador, pero en otros viene determinado por la naturaleza misma de la categoría.

Una investigación sobre dolores de espalda puede hacerse tomando cualquiera de las medidas o, incluso, todas a la vez; pero una investigación sobre el suicidio es difícil que tome otra medida que no sea la ocurrencia. Este hecho es muy importante a la hora de elegir los mejores indicadores para la observación de un determinado fenómeno.

Cuando sea posible habrá que elegir aquel indicador que mejor información nos reporte en relación con el objetivo de nuestra investigación.

5. La evaluación del proceso de observación

Hemos visto a lo largo de este capítulo que para llevar a cabo un proceso de investigación mediante observación, el investigador tiene que ir tomando una serie de decisiones en relación con la definición del problema, el procedimiento para observar, la elección de los participantes, las situaciones, los intervalos temporales. Tiene que decidir el modo de dejar constancia de lo observado. En caso de que lo considere necesario, tiene que construir su propio código de observación. Debe pensar en cuáles pueden ser los mejores indicadores para observar y el tipo de medida al que darán lugar...

Y todo ello de una manera tal que haga posible que cualquier otro investigador pueda replicar su trabajo. Muy probablemente pensará que es una "ardua tarea". Ciertamente no es fácil llevar a cabo este proceso sin que surjan problemas que puedan dar al traste con su objetivo. ¿Tiene el investigador alguna forma de saber si el proceso discurre correctamente? ¿Tiene algún modo de evaluarlo?

En toda tradición metodológica se han desarrollado procedimientos que hoy podríamos denominar de *control de calidad*. Son acciones que inserta el investigador en el proceso para comprobar que se está desarrollando dentro de unos parámetros. Fíjese en que estamos hablando sobre el proceso, no sobre el resultado. Es decir, nos planteamos si el modo en que se están recogiendo los datos de observación es adecuado, no si esos datos refutan o no las hipótesis de la investigación. Si el proceso está mal hecho, nada podremos concluir a partir de los datos que mediante él obtengamos. De los diferentes parámetros de control desarrollados en Psicología y disciplinas afines, aquí vamos a presentar dos de los más clásicos: la *fiabilidad* y la *validez*.

5.1. Fiabilidad: acuerdo entre observadores

Piense en algunos de los exámenes que ha hecho a lo largo de su vida de estudiante. Piense en un examen de preguntas abiertas para evaluar los conocimientos que hasta ahora ha trabajado dentro de este capítulo y el anterior. Imagine que uno de nosotros, los autores, el profesor León, es más "blando" a la hora de corregir, mientras que el otro, el profesor Montero, es más "duro". Bien, ¿cree que su examen obtendría la misma nota independientemente de quién lo calificara? Esperamos que su respuesta haya sido negativa. Piense ahora en un examen con preguntas cerradas con alternativas donde sólo una es correcta. Con esta modalidad de examen la respuesta a la pregunta anterior pasa a ser positiva: una vez conocida la lista de respuestas correctas, su nota sería la misma fuera quien fuera el corrector. La primera modalidad de examen es menos fiable que la segunda. La medida que nos provee de sus conocimientos merece menos confianza.

La *fiabilidad* hace referencia al grado en que podemos confiar en una medida.

Nadie confiaría en un procedimiento de observación que diera lugar a diferentes resultados en función de quién fuera el observador. Recuerde que estamos hablando de observación científica, no de una estimación política de la coyuntura socioeconómica actual. En términos que puedan estudiarse, un sistema de observación tiene que producir alto grado de *acuerdo entre observadores*. Es decir, dos observadores que presencian lo mismo tienen que estar de acuerdo en el producto de su observación.

Veamos un ejemplo. Piense en algunas decisiones polémicas de los árbitros en diferentes modalidades deportivas. Mejor aún, piense en el caso concreto del deporte de la canasta. El partido está a punto de acabar. El equipo de casa va dos puntos por debajo. Hay tiempo sólo para una posesión de balón. El entrenador pide tiempo muerto. La táctica está clara: preparan un tiro de tres para ganar el partido apurando la posesión al máximo para que no pueda haber réplica por parte de los rivales. Así se hace. Los segundos transcurren y, por fin, el balón llega a manos del especialista que ha conseguido una posición diáfana. Se eleva, lanza y... ¡canasta! ¡El partido está ganado! Pero... ¡la han dado como canasta de dos puntos! ¡El escándalo es monumental!...

Podríamos seguir esta crónica deportiva improvisada. Pero basta con hacerle caer en la cuenta de que la categoría "canasta de tres puntos", registrada *in vivo*, puede producir un cierto grado de desacuerdo entre observadores, mientras que la de "canasta de tiro libre", no. Podemos decir, por tanto, que una categoría es más fiable que otra. ¿Cuánto? Para ello existen índices que permiten dar un valor numérico a la fiabilidad así entendida.

El grado de acuerdo suele calcularse como *porcentaje de acuerdo*. Es decir, porcentaje de veces que los dos observadores registran lo mismo sobre el total de registros realizados entre ambos.

En el ejemplo baloncestístico, porcentaje de veces que dos árbitros expertos categorizan del mismo modo tiros realizados en el entorno de la línea de 6,25. Este índice de acuerdo no tiene en cuenta que parte del mismo puede deberse al azar. Cohen (1960) resolvió este problema proponiendo el uso de su coeficiente *Kappa*. Ambos índices –porcentaje de acuerdo y coeficiente *Kappa*– son fácilmente calculables.

Es recomendable que un código de observación –tanto en su conjunto como en cada una de sus categorías– obtenga, al menos, un 80% de acuerdo entre observadores. Este porcentaje podrá ser un poco inferior si se ha corregido mediante el coeficiente *Kappa*. Por debajo del 70% es claramente inaceptable. Lógicamente, es tanto más recomendable cuanto más se acerque al 100%.

5.2. Validez: de contenido, de constructo y referida al criterio

Volvamos a las dos modalidades de examen. No sabemos cuál es su opinión al respecto, pero muchos de nuestros estudiantes se quejan de que la modalidad de preguntas cerradas con alternativas no recoge bien lo que han aprendido. Piensan que otro tipo de examen reflejaría mejor sus conocimientos de la asignatura. Aunque no lo sepan, están cuestionando la validez del procedimiento de evaluación.

La *validez* de la observación hace referencia al grado en que se observa lo que se pretendía observar.

En nuestro ejemplo, la validez reclamada refleja el grado en que el examen mide los conocimientos que pretende medir. Lo que nuestros estudiantes reclaman es que el procedimiento para evaluarlos sea más válido. Advierta que achacan falta de validez a un procedimiento que, anteriormente, vimos que era el más fiable. Eso quiere decir que ambos criterios de calidad son independientes, aportan distinta información. Por tanto, deben exigirse ambos.

Todo proceso de observación tiene que ser a la vez fiable y válido.

La validez de un código de observación puede establecerse de diferentes modos. Nosotros nos vamos a referir a tres de ellos: validez de contenido, validez de constructo y validez referida al criterio. A continuación puede ampliar las modalidades de validez que se han propuesto.

Variantes del concepto de validez postuladas por los investigadores en Psicología

Cone (1978), en una obra ya clásica, recoge hasta siete variantes del concepto de validez aplicables según el tipo de investigación que se esté llevando a cabo. Le presentamos cada uno de ellos seguido de una brevísima definición del mismo. Recuerde que el objetivo de este cuadro es, simplemente, que tome conciencia de que existen más variantes de las que le presentamos. Nada más. No vaya a ser que le dé un mareo...

- **Validez aparente:** el procedimiento de recogida de datos sirve al objetivo para el que se desea utilizar.
- **Validez de constructo:** los indicadores para observar el constructo son un buen conjunto porque se dan las relaciones esperadas por la teoría entre ellos y otras variables a los que se ligan de forma deductiva.
- **Validez orientada al criterio:**
Concurrente (cuando la medida del criterio está presente en el momento de la observación).
Predictiva (cuando es necesario esperar a que aparezca la medida del criterio).
- **Validez convergente:** los indicadores observados correlacionan mucho con otras medidas que se sabe que guardan estrecha relación con la variable que pretenden medir.
- **Validez discriminante:** los indicadores observados no correlacionan con otras medidas que se sabe que son independientes de la variable que pretenden medir.
- **Validez discriminativa:** si los indicadores dan lugar a puntuaciones extremas en los grupos que, de antemano, se sabe que poseen valores extremos en la variable que pretenden medir.

Veamos con detenimiento cada una de ellas. Una vez más, el baloncesto será nuestro cómplice.

La *validez de contenido* de un código de observación hace referencia al grado en que el conjunto de categorías elaborado es una muestra representativa del universo de aspectos potencialmente observables.

¿Representan bien las categorías de "puntos anotados" –con sus tres modalidades–, "asistencias", "rebotes en ataque" y "robos de balón" a todas las posibles aportaciones de un jugador a la capacidad atacante de su equipo? Si le parece que sí, piense que el conjunto de categorías mencionado tiene una buena validez de contenido. Si cree que habría que añadir las "pérdidas de balón" y las "faltas en ataque", entonces está cuestionando la validez de contenido del conjunto propuesto. Recuerde que éste es uno de los modos posibles de expresar que ese código permite observar lo que pretendía: los aspectos relativos a la capacidad ofensiva de un jugador.

En el capítulo anterior le presentamos el concepto de *constructo*. Recuerde que un constructo es una variable teórica no directamente observable.

Cuando lo que pretendemos observar mediante un determinado código es un constructo, llegamos a una aparente contradicción: ¿cómo saber si hemos observado lo que pretendíamos cuando no lo podemos observar? No piense que le estamos gastando una broma. Queremos, simplemente, que se percate del adverbio *directamente*. Es decir, cuando utilizamos un conjunto de indicadores para observar *indirectamente* un constructo, podemos preguntarnos si ese conjunto nos permite observar lo que pretendíamos. Dicho de otro modo, podemos preguntarnos sobre su *validez de constructo*.

La *validez de constructo* es el grado en que un conjunto de indicadores acumulan evidencia a favor de la variable no directamente observable.

Siguiendo con el baloncesto, la estatura de un jugador es algo directamente observable; su rapidez, también. Pero su *capacidad ofensiva*... La *capacidad ofensiva* es un constructo. Es algo no directamente observable. Utilizamos una serie de indicadores que sí son directamente accesibles al observador para inferir la *capacidad ofensiva* de un jugador. Según definamos el constructo, el conjunto de indicadores observables necesarios para inferirlo será mayor o menor. Si consideramos que un jugador es buen atacante cuando anota puntos para su equipo, bastarán las tres categorías de anotación para inferir su calidad. Si pensamos que la *capacidad ofensiva* de un jugador tiene que ver también con que participe en las jugadas anotadoras, entonces incluiremos "asistencias", "rebotes ofensivos", "robos de balón"... El conjunto de las categorías elegido tendrá una buena validez de constructo si realmente informa sobre lo que la teoría del baloncesto considera que es un buen atacante.

La tercera variante del concepto de validez aplicado a la observación es la *validez referida al criterio*. Hace referencia al grado en que un código de observación detecta variaciones en aquello que pretende observar.

Cuando hablábamos del nivel de detalle relativo a la modalidad de anotación veíamos que no podríamos distinguir entre dos tipos de jugadores atacantes si no manejábamos un nivel superior de detalle. La categoría gruesa, "puntos anotados", sin más especificaciones, hacía que el código no permitiera recoger variaciones que resultan relevantes para elegir al mejor atacante en situación de tiro exterior a la zona de defensa del equipo contrario. La categoría así formulada resta validez al código cuando se orienta –se usa– para ese criterio: "mejor atacante exterior". Cuando incluimos las tres modalidades de puntos, ganamos sensibilidad, ganamos capacidad de detección de diferencias entre cualidades de atacante; en suma, ganamos validez referida a este criterio.

Todos estos controles de calidad que implican las preguntas sobre la fiabilidad y la validez de un código o de un proceso de observación suelen hacerse empíricamente. Recuerde el último paso en el procedimiento para elaborar un código. Era necesario depurar las categorías. Calcular el grado de acuerdo entre observadores o preguntarse si como conjunto son una muestra representativa es algo que hacemos después de un periodo de observación.

Existen, sin embargo, algunas fuentes potenciales de error que conviene conocer antes de llevar a cabo cualquier proceso observacional. Presentamos dos que consideramos de gran importancia y de fácil asimilación. Una proviene del observador y la otra del observado.

Cuando un observador se ha tomado la molestia de ir a recoger datos para poner a prueba alguna hipótesis o responder alguna pregunta, lo lógico es que tenga alguna expectativa sobre qué curso de los acontecimientos encajaría mejor con su hipótesis o pregunta. Esta expectativa puede convertirse en una fuente de error al hacer categorizar de forma errónea algún elemento del fenómeno. En los casos dudosos, las canastas de tres puntos se catalogan, o no, en función de qué equipo sea el anotador. En el relato que hacíamos del final del partido, los espectadores y los miembros del equipo atacante han visto canasta de tres. Los árbitros –y seguro que los visitantes– han visto canasta de dos. Las diferentes apreciaciones no son un problema del cinismo del observador. El observador no hace trampa al observar lo que observa, simplemente sus expectativas le llevan a error.

La manera de controlar este posible sesgo es poner a observar alguien que sea ajeno a los objetivos de la investigación. Esto suele hacerse entrenando a estudiantes de últimos años o a investigadores en formación a los que se informa de los objetivos, al final del proceso. Mientras dura la observación decimos que son *observadores ciegos*¹². Por eso los árbitros no son de ningún equipo. O, al menos, eso se supone...

La otra fuente de problemas proviene del sujeto al que observamos. Está demostrado que los seres vivos reaccionamos al hecho de ser observados. Cuando notamos la presencia de alguien que observa, dejamos de comportarnos con naturalidad y, por tanto, ya no se nos observa en las condiciones en que se pretendía. Este fenómeno se denomina *reactividad*. El jugador novato, recién ascendido, que rinde mal al ver tanta gente observándolo en la cancha, sería un ejemplo de reactividad. Hay dos maneras de obviarla: mediante engaño (no descubrir la cámara hasta el final) o mediante habituación. Por su-

12. En el capítulo III ("Metodología experimental") , cuando tratemos los experimentos con una sola variable, tendrá ocasión de comprobar que hay elementos que no pasan por la percepción consciente de los jueces deportivos que influyen sobre las decisiones que éstos toman cuando arbitran.

puesto, hay situaciones que pertenecen a la esfera privada de las personas y que no podemos observar ni aunque después les avisemos de que lo hicimos (consulte el Apéndice C).

Capítulo III

Metodología experimental

1. Experimentos con una variable independiente; diferentes sujetos

1.1. Introducción: ¿qué es experimentar en Psicología?

Presentación de un caso

Suponga que queremos investigar cómo se comportan las personas en situación de presión de tiempo; en concreto, a la hora de ir al trabajo por la mañana. Podemos preparar un cuestionario en el que se pregunte por su comportamiento y por el de sus semejantes y seleccionar una muestra representativa para ser interrogada (*metodología de encuestas*). Suponga que dudamos de la sinceridad de los encuestados, dado que la deseabilidad social les puede hacer aparecer más civilizados de lo que realmente son.

Podemos adoptar otra estrategia: nos apostamos en la salida del metro de una gran ciudad a primera hora de la mañana; anotamos todas las conductas de las personas que aparecen durante un periodo de cinco minutos (*metodología de observación*). Sin embargo, nos surgen dudas: algunos no tendrán prisa, incluso llegan con tiempo para tomar un café antes del trabajo; otros no van a trabajar, sino que vuelven de fiesta; algunos llegan media hora tarde, mientras que otros llegan sólo un par de minutos tarde; en esos cinco minutos, unos piden un taxi, otros corren, otros cruzan por un semáforo, otros son asaltados por un indigente...

La lista de circunstancias que afectan a la conducta, además de la prisa por llegar al trabajo, es grande (*variables extrañas*). Dudamos de la capacidad de obtener datos consistentes. Necesitamos crear unas condiciones producidas por nosotros (*variable independiente*) para estar seguros de lo que está pasando (*variable dependiente*) y poder interpretarlo.

Por ejemplo, Darley y Bateson (1973) crearon dos *condiciones*. En una, los participantes eran forzados a llegar tarde, y en la otra, no. En medio del proceso aparecía un indigente solicitando ayuda. ¿Tenía algún efecto la prisa sobre la solidaridad hacia el indigente? El grupo que tenía prisa ayudó en un 10% de los casos, mientras que el grupo que no tenía prisa ayudó en un 63% de los casos. Como los investigadores hicieron que ambos grupos fueran iguales excepto en la prisa, pudieron concluir que la prisa hacía que la gente fuera menos solidaria.

A esta forma de investigar la llamamos *metodología experimental*: producimos el fenómeno, controlamos las circunstancias y decidimos quién estará bajo qué condiciones. Fíjese en lo difícil que habría sido concluir que la prisa estaba relacionada negativamente con la conducta de ayuda mediante la encuesta; muchos de nosotros, de haber sido encuestados quizá dijéramos de buena fe que ayudaríamos, pero la realidad mostraría otra cosa; también hubiera sido difícil mediante observación decir que los que no ayudaban eran los que tenían más prisa, ya que no sabríamos si se debía a otra circunstancia.

Mediante la experimentación podemos contrastar que una variable es la causa de otra.

Veamos ahora con detenimiento cómo se desarrolla un experimento en Psicología. El blanco y el negro han simbolizado desde antiguo el bien y el mal en muchas culturas. Por ejemplo, los justos que gozan en presencia de Dios, tal como los describe San Juan en el Apocalipsis (Juan, Ap, 7, 9), llevan vestiduras blancas. Darth Vader, el representante del lado oscuro en la saga de *La guerra de las galaxias*, viste de negro riguroso. Mozart, en la *Flauta Mágica*, contrapone la reina de Noche y el sacerdote del Sol; los caballos de Ben Hur son blancos, mientras que los de Mesala son negros. Incluso Rowling (1999), en su obra del superhéroe infantil Harry Potter, encasilla al malo Voldemort como el representante del lado oscuro. ¿Hasta qué punto los colores blanco y negro condicionan nuestra percepción, o se trata sólo de signos culturales?

Para probarlo, Frank y Gilovich (1988) realizaron la investigación siguiente. Primero analizaron las estadísticas: a los equipos de fútbol americano y de hockey con camiseta negra les pitaban más faltas que a los que llevaban camiseta blanca. Este tipo de dato –descriptivo– no les aseguraba que el responsable fuera el color de la camiseta. Quizá fuera otra la causa. Es incluso posible que los equipos que se perciben a sí mismos como más agresivos elijan el color negro.

Como hemos comentado en las metodologías con encuestas, incluso con altas correlaciones, no podemos asegurar la causación entre dos variables. Los autores necesitaban reproducir el fenómeno bajo su entero control, de forma que otras explicaciones distintas de la suya pudieran ser descartadas. Necesitaban planear un experimento.

Su hipótesis fue que si un jugador de fútbol americano lleva una camiseta negra, entonces se le pitarán más faltas que cuando lleva una blanca. Un experimento de campo sería muy costoso: por ejemplo, hacer que un equipo jugara la mitad de sus partidos de blanco y la otra mitad de negro... Podría ser que algunos sintieran que su honor se ponía en entredicho al ponerse la elástica blanca.

Su proyecto fue más sencillo: primero recopilaron vídeos con jugadas dudosas o sancionadas, en las que los supuestos infractores vistieran unas veces (50%) de blanco, y otras (50%) de negro. Después, un grupo de árbitros profesionales, discutiendo entre ellos y con moviola, valoraría si había falta o no. El resultado final sería un vídeo donde "los de blanco" cometían el mismo porcentaje de faltas que "los de negro".

Pasemos ahora a los participantes. ¿Quiénes podían ser jueces de fútbol?

Cualquiera no servía. En primer lugar, tenían que conocer correctamente las reglas del juego. En segundo lugar, debían verlo habitualmente. En tercer lugar, no podían ser seguidores –ni "adversarios", se entiende– de ninguno de los equipos que aparecían en las grabaciones.

Sus edades no debían ser extremas, de forma que establecieron un rango entre 25 y 55 años. Mantuvieron un discreto 15% de chicas, similar al porcentaje de asistentes a los campos de fútbol.

Como la tarea consistiría en ver jugadas en un monitor, seguro que descartaron algún candidato con mala visión. Los participantes –60– se escogieron entre los que contestaron a una solicitud por escrito, repartida en un estadio durante varios partidos, en una gran ciudad del este de Estados Unidos.

Para estar más seguros de que era el color de la camiseta lo que determinaba el mayor número de sanciones de faltas, diseñaron la estrategia siguiente. En una copia con las jugadas conflictivas modificaron digitalmente el color de todas las camisetas, de manera que todos los equipos supuestamente infractores aparecían de gris. De esta forma podían ver cuántas faltas les "pitaban los jueces", sin estar afectados por el color de las camisetas; a esta cinta en gris, la llamaremos *control*. En la otra cinta aparecían los colores blanco y negro; a esta cinta la llamaremos *experimental*.

Ya sólo faltaba que la mitad de los jueces vieran –y pitaran– las jugadas de la cinta control y la otra mitad las de la experimental. Por supuesto, los dos grupos de jueces debían ser equivalentes, no podía ser que unos fueran

más estrictos que otros. Para conseguirlo hicieron una lista de todos por orden alfabético y después les pusieron un número de orden. Sacaron un número al azar entre 1 y 60. El juez con ese número fue asignado al vídeo control, el juez con el número siguiente al experimental, y así sucesivamente hasta agotar la totalidad de los jueces.

Si quiere saber cómo se lleva a cabo una *asignación aleatoria*, puede practicar haciendo la actividad 3.1., en la página web.

En la página web se dispone de una actividad complementaria (capítulo III, apartado 1.1).

Veamos cómo están las cosas hasta este momento: tenemos dos cintas de vídeo que son idénticas, excepto el color de las camisetas de los infractores; tenemos dos grupos igualados de jueces dispuestos a valorar cada cinta. Si el color de la camiseta no influye en su decisión de pitar falta, en ambas cintas el equipo "negro" recibirá las mismas faltas que el equipo "blanco".

Los resultados fueron los siguientes:

- a) en la *condición* de control ambos equipos recibieron el mismo número de faltas;
- b) en la *condición experimental* el equipo con camiseta negra recibió el doble de faltas que el equipo con camiseta blanca (Frank y Gilovich, 1988).

1.2. Diseño con grupos aleatorios: la hipótesis; los participantes y su asignación; el diseño, las variables y su operativización; los resultados y la discusión

En el experimento de Frank y Gilovich (1988), se utilizó un plan que consistió en disponer de dos grupos de participantes que realizarían una tarea bajo dos condiciones distintas. Es decir, dos grupos de jueces que observarían el mismo vídeo y atribuirían faltas, bajo dos condiciones distintas, como eran los distintos colores de las camisetas. El grupo de jueces que veía a los jugadores vestidos de gris formaba el *grupo control*. El grupo de jueces que analizaba las faltas de jugadores con camiseta blanca o negra formaba el *grupo experimental*. Veamos cuáles podían ser las razones para desarrollar esta estrategia.

Primera, los dos grupos de jueces debían ser distintos, porque se quería mantener el mismo material estimular (las mismas jugadas conflictivas) y, por lo tanto, los mismos jueces no podían ver dos veces la misma cinta; esto alteraría las respuestas de la segunda vez.

Segunda, fueron dos los grupos, porque había dos condiciones experimentales (o dos niveles de la variable independiente), pero podrían haber hecho un plan con tres o con cuatro grupos (podían haber contrastado si el rojo o el verde tenían algún efecto diferencial). Cuando se utilizan sólo dos grupos y uno de ellos está en una condición en la que recibe un nivel nulo de la variable independiente, a éste se le llama *grupo control*; al otro grupo, que recibe un valor cualquiera de la variable independiente, se le llama *grupo experimental*. El grupo control sirve como nivel de partida para valorar la cuantía del efecto del grupo experimental; en este contexto, no asumimos que los de gris no van a realizar ninguna falta, sino saber cuántas faltas se ven en las jugadas, cuando el color de la camiseta no afecta nada al observador.

Tercera, los dos grupos fueron generados mediante la acción del investigador. Ninguna otra circunstancia externa determinó la pertenencia al grupo experimental o al grupo control. El investigador tuvo en su mano decir quién iba a cada grupo. La forma en que dividió a los participantes (asignación aleatoria) es la que da nombre al diseño: *diseño de grupos aleatorios*. En el diseño con grupos aleatorios el investigador genera tantos grupos como niveles tiene la variable independiente y mediante un procedimiento basado en el azar reparte a los participantes entre los grupos. El azar consigue que las variables extrañas asociadas a las diferencias individuales queden repartidas y, por lo tanto, que los grupos sean equivalentes (o, en expresión sinónima, que estén equilibrados). Un buen plan de investigación con grupos de participantes distintos debe asegurar que los grupos reaccionarán de la misma manera ante la misma condición experimental. Asegurado esto, si reaccionan de forma distinta, deduciremos que se debe a los distintos niveles de la variable independiente. Para significar que distintos participantes pasarán por los distintos niveles de la variable independiente, también se utiliza como abreviación la palabra *inter*. Así, se dice de una variable que es *inter* o *intersujetos*; en reciprocidad, cuando son los mismos participantes los que pasan por todas las condiciones de la variable independiente, se utiliza la expresión *intra*, y se dice que tal variable es *intra* o *intrasujetos*. Así, en este experimento decimos que la variable color de las camisetas es intersujetos.

En todo experimento hay explícita una hipótesis de relación causal. En nuestro caso era:

"si el color negro en la camiseta de los jugadores hace que su conducta se perciba como más beligerante en una situación dudosa, entonces cuando esos mismos jugadores, en las mismas jugadas, lleven una camiseta de color gris recibirán menos sanciones que cuando la lleven negra."

Puede observar dos cosas en esta hipótesis:

En primer lugar, tiene una estructura "si..., entonces..." que encadena la variable independiente con la variable dependiente y establece con claridad qué es lo que debe suceder en el caso de que la relación que se establece en el antecedente sea cierta.

En segundo lugar, está redactada con un elevado grado de meticulosidad; decimos que está operacionalizada u *operativizada*: está concretada de manera que lo que se expresa se puede observar y registrar.

En la exposición, hemos utilizado el término *variable* con diferentes adjetivos y ahora conviene que lo sistematicemos.

El color de las camisetas era la variable independiente; que hipotéticamente produce el efecto. Sus niveles eran las formas en que se presentaba: primer nivel, negro/blanco; segundo nivel, gris/gris.

El investigador tenía en sus manos la capacidad de presentar estos niveles y de hacer que los participantes que él quisiera se encontraran bajo un nivel u otro (aunque ya sabemos que confió la tarea al azar). Por este motivo, se dice que el investigador tiene la capacidad de manipular la variable independiente.

La calificación de falta en el juego del vídeo era la variable dependiente. Aquella conducta de los participantes estaba influida en potencia por la independiente. Para que esté correctamente operativizada, es preciso especificar la manera de medirla. En nuestro caso se trata del número de faltas adjudicadas al equipo cuando lleva la camiseta negra y cuando lleva la gris.

En el experimento había otras variables sobre las que el investigador no hipotetizaba nada en relación con la dependiente. Sin embargo, ante la sospecha de su influencia las tenía en cuenta y las controlaba. Así, las variables controladas eran las siguientes:

- *El nivel de conocimientos técnicos*. Controlada mediante igualación, ya que todos entraban en la misma categoría "con conocimientos adecuados".
- *La asistencia a partidos*. Controlada mediante igualación. Todos asistían regularmente a ver partidos.

- *La edad*. Controlada por equilibrado. Es decir, al haberse repartido por asignación aleatoria a los participantes en dos grupos, habrá personas de todas las edades en las dos condiciones.
- *El género*. Controlada por equilibrado. La asignación aleatoria habrá repartido hombres y mujeres de forma no sesgada a los dos grupos.
- *La pasión y la animadversión*. Controlada por eliminación. Aquellos seguidores –o adversarios– de los equipos filmados quedaron fuera de la investigación.
- *La capacidad de visión*. Controlada por eliminación. Aquellos candidatos con dificultades de visión no participaron; el resto, con sus diferencias naturales, veía con normalidad.

¿No hay más variables que pudieran influir en la dependiente? Seguramente, pero en algún sitio hay que cortar. Lo que suelen hacer los investigadores es leer antes en otros trabajos publicados qué variables influyen y en consecuencia controlarlas. Que no se conozcan no quiere decir que no estén presentes. ¿Qué pasa, pues, con ellas? Actuar, actúan; y la prueba es que no todos los participantes bajo la misma condición se comportan de la misma manera. Justamente, estas diferencias son la prueba de que hay otras variables que afectan a la dependiente. Como no se conocen, reciben el nombre de *variables extrañas*.

Dentro de las variables extrañas existe una clase particularmente perniciosa: las *enmascaradas*¹³.

Cambiamos de contexto de investigación. Suponga que leemos en un estudio que se ha comprobado que las familias españolas que han decidido vivir en casas de madera gozan de mayor salud que las que viven en casas de ladrillos. Como explicación se aduce su mejor aislamiento y su capacidad de "respirar". Existe una posible variable enmascarada: que las familias que viven en casas de madera tengan mayor poder adquisitivo. Si esto es así, el poder adquisitivo cambia al mismo tiempo que cambian los niveles de la variable independiente y es una variable que puede explicar las diferencias en la dependiente. Una vez detectada la posible variable enmascarada, el investigador debe controlarla. En este caso, eligiendo familias en casas de ladrillo con similar poder adquisitivo al de las familias que viven en casas de madera.

Cuando leemos un informe de investigación no consideramos aceptable que nos digan cosas como "el grupo que vio el vídeo en colores fue menos deportivo", o, en otro contexto, "los enfermos que no siguieron tal terapia consiguieron levantar su plano vital". Cuando leemos un informe de investigación

13. En inglés *confounded variables*, lo cual ha dado pie a que se haya traducido de manera demasiado literal como variables confundidas, expresión que evoca una semántica distinta a su concepto.

esperamos encontrar una sección de "resultados" donde se exprese de forma numérica cuáles han sido los valores de la variable dependiente bajo cada condición de la independiente.

Cuando se ha utilizado un diseño con grupos, hay que resumir la información de cada grupo mediante un índice de tendencia central. El más utilizado es la media.

En el caso que hemos relatado, se informó que el número de faltas pitadas al equipo con camisetas de negro era el doble que el número pitado al equipo con camisetas blancas (en el original aparecen los valores concretos).

Recuerde que las hipótesis no se demuestran ni se confirman ni se verifican, sino que se mantienen de manera provisional. Los resultados se discuten en relación con la hipótesis planteada, para inferir si puede mantenerse provisionalmente o rechazarse. Asimismo, en la fase de discusión deben relacionarse los hallazgos presentes con los de otros autores, de forma que el conocimiento no se parcele en exceso y permita establecer reglas lo más generales posible.

Las cualidades de la investigación con experimentos són: control, fiabilidad, sensibilidad, validez interna y validez externa.

1)Control

La primera cualidad que distingue la metodología experimental es el *control*. Veamos en qué indicios del experimento anterior podemos descubrir acciones de control:

1) "Videos con jugadas dudosas o sancionadas". De esta forma la atención de los jueces se centraba sobre el problema y no tenían que ver todo el partido.

2) "Los supuestos infractores vestían unas veces (50%) de blanco y otras (50%) de negro". El resultado final sería un vídeo donde "los de blanco" cometían el mismo porcentaje de faltas que "los de negro". De esta forma, si el color no afecta, se espera que los jueces "vean" las mismas faltas en "los de blanco" que en "los de negro".

3) "Los jueces tenían que conocer correctamente las reglas del juego". Para esta investigación no servía cualquiera. Además de ser aficionado, debía tener un buen conocimiento técnico.

4) "Debían verlo [el fútbol] habitualmente". Esta condición complementa a la anterior; no sólo debían conocer las reglas, sino ver partidos regularmente para poder ser buenos jueces.

5) "No podían ser seguidores –ni adversarios–". Con esta exigencia controlaban una de las variables que más prejuicios ocasiona en la valoración de faltas en el deporte de equipos.

6) "Sus edades (...) entre 25 y 55". Con esta franja de edad los investigadores tratan de representar al público que habitualmente sigue los partidos.

7) "Chicas, (...) un 15%, similar al porcentaje de asistentes a los campos de fútbol". No sabemos si las mujeres son más o menos estrictas que los hombres al considerar las faltas. Los investigadores optaron por una representación similar a la del público que asiste a los partidos.

8) "Descartaron algún candidato con mala visión". Puesto que la tarea consistía en valorar lo que se veía en un monitor, había que ver bien.

9) "Los dos grupos de jueces debían ser equivalentes". Con esta medida de control se aseguraron que los dos grupos estarían afectados por la mismas variables extrañas cuando juzgaran los vídeos. Esta técnica de control –asignación aleatoria– será expuesta en el próximo apartado.

Las acciones 1 y 2 fueron encaminadas a obtener una tarea que permitiera generar el fenómeno deseado, en condiciones de "laboratorio". Junto a estas indicaciones, había otras como el tamaño del monitor y sus características, la distancia a la que estaban los jueces, la hora del día... todas ellas medidas de control con cuya especificación otros investigadores podrán replicar el experimento.

Las medidas de control 3 a 8 van encaminadas a que los participantes de esta investigación sean los adecuados para que pueda producirse el fenómeno que se estudia. (Nosotros difícilmente serviríamos para analizar jugadas dudosas de fútbol americano.)

La medida de control 9 va encaminada a que la única diferencia entre las dos condiciones de visión del vídeo fuera la ausencia –todos de gris– o presencia del color –de blanco y de negro–. Sorteando los jueces a cada condición nos aseguramos de que un grupo de jueces no difiere del otro.

¿Cree que podíamos haber adoptado estas medidas de control con las metodologías descriptivas?

El control es una de las características esenciales de la metodología experimental.

2) Fiabilidad

¿Fue *fiable* este experimento? Cuantas más medidas de control tenga un experimento es más probable que sea fiable. Sin embargo, hasta que no se repita no lo sabremos. Como ejemplo reciente de poca fiabilidad podemos citar los casos de los laboratorios que obtenían fusiones en frío y luego no podían repetirlos.

La *fiabilidad* es el grado en el que se estima que, al realizar de nuevo un experimento, obtendremos los mismos resultados.

Todos los procesos que hacen aportaciones relevantes para la ciencia son repetidos en numerosas ocasiones, ya que, a veces, algunos resultados deslumbrantes no se vuelven a obtener. Cuando al repetir un experimento no se obtienen los mismos resultados, eso significa que alguna variable no se controló bien, o que existe alguna variable desconocida que interviene de forma diferente en cada repetición. ¿Qué cree que hubiera pasado si en lugar de ver un vídeo con una selección de jugadas dudosas hubieran visto un partido completo? Lo más probable es que en todo el partido aparecieran, como mucho, media docena de jugadas dudosas. Con este reducido número es probable que no se pusieran de manifiesto las diferencias entre el vídeo en gris y el coloreado.

3) Sensibilidad

Cuando un experimento no tiene la capacidad para producir el fenómeno en estudio, decimos que no tiene *sensibilidad*.

Esto puede ocurrir por múltiples causas; en nuestro caso, por un mal conocimiento técnico de los participantes o porque la toma de la cámara fue tal que no se apreciaba bien la responsabilidad en la acción.

4) Validez interna

Probablemente, la cualidad más identificativa de las metodologías experimentales es la validez interna. Lo primero que vamos a hacer es definirla con una sola palabra del lenguaje cotidiano para facilitar su conocimiento: *causación*.

Una investigación, y en este caso un experimento, tiene una adecuada validez interna si cumple tres requisitos: antecesión de la variable independiente, covariación entre la variable independiente y la dependiente, y capacidad para descartar explicaciones alternativas¹⁴.

14. Esta ordenación está hecha siguiendo la secuencia temporal en que se realiza la investigación.

a) **Antecesión**

¿Se produjo la manipulación del color de las camisetas antes de que se produjeran las atribuciones de faltas? La respuesta es que sí. Este requisito es fácil de cumplir, ya que, al tratarse de un experimento, la acción de la variable independiente está bajo el dominio de los investigadores y son ellos los que presentan la tarea de la forma adecuada. Primero aparece la variable independiente y después se registra la dependiente. Para que pueda decirse que existe causación entre dos variables debemos estar seguros de cuál antecede.

En la página web se dispone de una actividad complementaria (capítulo III, apartado 1.3).

El primer requisito para asegurar la validez interna es que la variable independiente se produzca antes que la dependiente.

b) **Covariación**

¿Qué ocurre con el número de faltas cuando un equipo lleva camiseta blanca o lleva camiseta negra? Respuesta: las faltas aumentan. Es decir, cuando se cambian los valores de la variable independiente cambian los valores de la dependiente. A este fenómeno, en estadística, se le llama *covariación*: cambiar simultáneamente los valores de dos variables.

El segundo requisito para que exista validez interna es que se dé covariación entre la variable independiente y la dependiente.

c) **Capacidad de descartar explicaciones alternativas**

¿Estamos seguros de que la única causa por la que se pitaron distinto número de faltas en las dos condiciones era el color de la camiseta? Recuerde: el equipo objetivo era visto en el mismo vídeo, en condiciones similares, por dos grupos de jueces similares, y la única diferencia de las dos sesiones era que en una iba de gris y en la otra de negro... Si la respuesta es afirmativa, entonces habremos cumplido la tercera condición para hablar de causación.

Para asegurarnos de este tercer requisito, además, deberemos estar en condiciones de rechazar otras explicaciones alternativas. Por ejemplo, si alguien dijera que "eso ocurrió porque en el grupo de jueces que pitó más faltas había más hombres", nosotros debemos justificar que la proporción de hombres fue similar en ambos grupos (cosa que se hizo mediante la asignación aleatoria). Si alguien dijera que "a lo mejor, el grupo de jueces que pitó menos faltas sabía menos de fútbol", tendremos que rechazarlo argumentando que la calificación de ambos grupos fue similar y que se aseguró mediante la asignación aleatoria.

Por lo tanto, podemos afirmar que se cumple el tercero, y último, de los requisitos para poder asegurar la causación y evaluar positivamente el grado de validez interna del experimento.

5) Validez externa

La mayor aportación de la metodología experimental es poder contrastar la existencia de causalidad entre una variable y el fenómeno que, supuestamente, produce.

¿Cree que pitar más faltas a los "de negro" es exclusivo de esta situación experimental o cree que se dará también en los campos de juego? Si piensa que ocurrirá lo mismo en los campos de juego es porque este experimento tiene capacidad de generalización, tiene *validez externa*. Suponga por un momento que las condiciones en que se realiza la visión del vídeo incluyen: cámara superlenta, moviola tantas veces como desee el espectador, trazado de líneas adicionales en la pantalla, discusión en grupo por parte de los participantes y consulta telefónica directa con un árbitro profesional. ¿Piensa ahora que los resultados de este segundo experimento podrían generalizarse? Seguramente pensará que no. Le faltaría validez externa.

A continuación presentamos otro ejemplo de un experimento con distintos participantes.

Caso 3.1.

¿Le pueden inducir a ver lo que no ha visto?

Algunos investigadores como Loftus, Miller y Burns (1978) han mostrado experimentalmente que nos pueden engañar haciéndonos recordar cosas que no hemos visto. Para probarlo llevaron a cabo varios experimentos del estilo del que se describe a continuación. Su hipótesis era:

"Si en un interrogatorio le inducen a pensar que cierto suceso ocurrió, posteriormente, al preguntarle por ese suceso es probable que crea que sucedió."

Cien estudiantes fueron divididos al azar para formar un grupo experimental y otro control. Tanto el grupo experimental como el control vio veinte diapositivas en torno a un atropello. En una de ellas aparecía un coche en un cruce junto a un poste que no tenía ninguna señal. Después de veinte minutos haciendo una tarea para distraerlos, se les pasó un cuestionario sobre las diapositivas.

El cuestionario fue igual para los dos grupos excepto una pregunta malévola respecto a la diapositiva del cruce.

A los participantes del grupo control se les pedía que dijeran si recordaban el color del coche que estaba en el cruce. A los participantes del grupo experimental se les pedía que dijeran si recordaban el color del coche que estaba "junto a la señal de ceda el paso". Nuevamente fueron entretenidos y un último cuestionario apareció. En éste, una pregunta para los dos grupos pedía recordar si en el cruce había un ceda el paso o nada.

El grupo control informó que recordaba un ceda el paso en un 2% de los casos. El grupo experimental informó que recordaba un ceda el paso en un 75% de los casos. Es decir, la pregunta anterior, que daba por supuesta la existencia de la señal, hizo que el 75% de los testigos creyeran de buena fe que habían *visto* una señal de ceda el paso. (Para más información sobre falsos recuerdos, puede consultar Diges, 1997).

Cuando las condiciones de los laboratorios se vuelven muy artificiales, se emplea también la expresión de *falta de validez ecológica*, que expresa las diferencias entre el laboratorio y el contexto natural.

Diseño de grupos aleatorios con una variable bloqueada

En la actividad 3.1 de la página web veíamos cómo mediante la asignación aleatoria de los participantes podíamos conseguir que las variables extrañas asociadas a las diferencias individuales quedaran equilibradas entre las condiciones experimentales. Ese "quedar equilibradas", o producir grupos equilibrados, se concretaba en que las variables extrañas se repartieran de "forma no sesgada". En concreto, el porcentaje de chicos en el grupo experimental fue del 40%, y en el control, del 60%. En una variable dependiente donde el género tenga un papel secundario, este reparto podemos considerarlo adecuado. Pero ¿qué ocurre si el género es una variable muy relevante respecto a la dependiente?

Por ejemplo, suponga que investigamos sobre trastornos en la conducta alimentaria de los adolescentes (Toro y Castro, 2000). Primero, elaboramos un cuestionario que identifica predictores de futuros desórdenes. Posteriormente, con el cuestionario desarrollado, se identifican veinte adolescentes de riesgo, con edades entre los trece y los dieciséis años, dentro de un gran centro de educación secundaria. De los veinte, catorce son chicas (70%).

La fase siguiente consiste en aplicar un programa para modificar actitudes y hábitos alimentarios. Se divide el grupo al azar en dos grupos de diez, para participar en el programa o en el grupo control –en espera–.

Finalmente, se comprueba que el programa produce cambios en el grupo experimental al compararlo con el de control. Sin embargo, al redactar las conclusiones nos damos cuenta de que, por "azar", estaba compuesto sólo por un 40% de chicas. ¿Estaríamos seguros de que produciría los mismos cambios si ese grupo hubiera sido compuesto por un riguroso 70%? No lo estaríamos.

Por lo tanto, en este caso, como el grupo experimental y el de control tienen diferentes porcentajes en cuanto al género de los participantes, no podemos desechar una explicación alternativa que dijera que las diferencias entre los grupos del experimento se debieron a la diferente composición de sus participantes.

En la página web se dispone de una actividad complementaria (capítulo III, apartado 1.2).

Si el género debe ser correctamente controlado debemos encontrarnos con porcentajes *iguales* de chicos y chicas en los grupos del experimento. No podemos confiar en que el azar repartirá de *forma exacta* un 70% a cada uno de los grupos, especialmente cuando se trata de grupos pequeños. Si mediante el azar no podemos asegurarlo, sí podemos conseguirlo mediante una adecuada estrategia. Veamos cómo se haría en este caso:

- 1) Dividimos a los participantes en dos bloques: uno de chicas (14) y el otro de chicos (6).
- 2) Repartimos a los componentes del bloque de las chicas –al azar– a cada una de las condiciones experimentales, de manera que siete irán al grupo experimental y otras siete al de control.
- 3) Repartimos a los componentes del bloque de los chicos –al azar– a cada una de las condiciones experimentales, de forma que tres irán al experimental y tres al de control.

Esta técnica de control recibe el nombre de *diseño de bloques aleatorios*. Recuerde que los bloques se hacen en una variable extraña, cuyo efecto sobre la dependiente deseamos controlar.

¿Qué variable elegir? Debe ser una variable que influya con seguridad sobre la dependiente y que en el experimento no va a ser manipulada. Puesto que la variable que se elige para bloquear es relevante para explicar las variaciones en la dependiente, tampoco conviene que todos los participantes estén igualados en valor. No hubiera sido bueno que los dos grupos fueran sólo de chicas. De hacerlo así, los participantes no representarían bien al resto de la población. Deseamos que se dé todo el rango natural de sus valores. Si, como en nuestro ejemplo, existen chicos con trastornos alimentarios, la variable género debe tener sus dos aspectos representados.

- En la variable extraña, objeto del control, elaboraremos tantos bloques como queramos –cuantos más bloques, más control.

- Los bloques pueden tener todos la misma medida o diferente, como en nuestro caso. Dependerá de la naturaleza de la variable.
- A continuación, aplicamos a cada bloque la técnica de asignación aleatoria para repartir a los integrantes del bloque en las condiciones experimentales; de este modo, el resto de las variables asociadas a las diferencias individuales quedarán controladas por equilibrado.

En definitiva, con la técnica de control de una variable extraña mediante bloques aleatorios en esa variable, conseguimos una mayor validez interna en el experimento, lo cual redundará en que sea más fácil poner de manifiesto la relación causal entre la independiente y la dependiente (adicionalmente, en el análisis de datos, puede eliminarse de la variancia del error la parte debida a los bloques, haciendo que el efecto debido a la VI sea mayor.)

2. Experimentos con una variable independiente; los mismos sujetos

2.1. Presentación de un caso

Le invitamos ahora a que proyecte con nosotros la réplica de un experimento clásico de Eckman, Levenson y Friesen (1983) sobre el papel de las expresiones faciales en las emociones. Uno de los hallazgos de esta investigación fue que "poner caras" activando los músculos implicados en la expresión de emociones –pero sin nombrarlas– produce algunas respuestas fisiológicas como si se experimentaran las emociones de verdad.

En concreto, vamos a contrastar si las expresiones faciales ira, miedo y alegría conllevan unas tasas cardíacas (TC) superiores a la tasa cardíaca inicial (antes de ejecutar las expresiones).

Nuestra variable independiente es la expresión facial con los tres niveles especificados. La forma de operativizarla será siguiendo las instrucciones para activar adecuadamente los músculos que forman cada expresión. Por ejemplo, para representar el miedo hay que juntar y elevar las cejas, elevar los párpados y estirar la boca horizontalmente. La variable dependiente la mediremos con un pulsómetro¹⁵.

¹⁵

Si seguimos la estrategia del diseño de grupos aleatorios necesitaremos tres grupos de participantes equivalentes en su TC y con habilidad similar para controlar sus músculos faciales. La TC es una medida fisiológica que, dentro de la normalidad, presenta grandes diferencias individuales, lo cual nos obligaría a formar grupos grandes de participantes para que la variabilidad de la TC, antes de empezar, quede equilibrada entre las condiciones experimentales. Adicionalmente, encontrar grupos de personas –actores– con similares habilidades obligaría a evaluarlos previamente. Deberíamos, además, controlar todas las variables que pueden afectar a la TC: edad, género, práctica deportiva habitual, posibles enfermedades, etc. En definitiva, este diseño para este propósito no es el mejor.

Lo adecuado aquí es utilizar el mismo grupo de actores para representar cada una de las tres emociones. De esta forma los tres "grupos" son absolutamente equivalentes: de hecho son el mismo grupo en tres situaciones. Una vez conseguidos los participantes (por ejemplo, en el último curso de una escuela de arte dramático) nos aseguraremos de que son capaces de ejecutar correctamente las instrucciones. Tendremos especial cuidado en no decirles el propósito de la investigación ni nombrar ninguna de las tres emociones, para evitar que utilicen imágenes o recuerdos para ayudarse en la expresión. Recuerde que el objetivo es ver el efecto de la configuración muscular, sin etiqueta emocional.

¿Por qué emoción empezamos? ¿Le parece que empezemos por ira, luego miedo y por último alegría? (de esta forma acabarán el experimento más "contentos"), o ¿ponemos la alegría en medio para que no haya tanta expresión negativa junta? ¿Le parece que da igual o intuye que deberíamos tener en cuenta el orden? Si se siente inclinado hacia la última opción, ha acertado. En esta investigación, es importante controlar el orden en que se ejecutan los tres niveles de la variable independiente. Es posible que cueste más representar ira después de alegría que representarla después de miedo; aunque esta tarea es muy corta, quizá un ligero cansancio –en la última representación– afecte a la TC; o un ligero nerviosismo en la primera. Es más, deberíamos asegurarnos de que cuando se comienza una expresión la TC ha vuelto al mismo nivel de antes de la primera expresión, ya que la TC tarda unos minutos en recuperarse después de pasado el estímulo que la incrementó. Teniendo en cuenta lo anterior, decidimos que las tres expresiones sean representadas en estos tres órdenes elegidos, de entre los posibles:

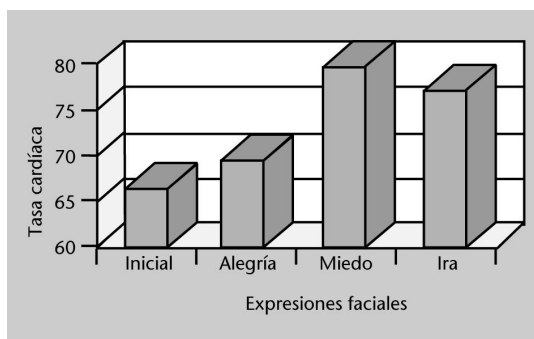
15. El pulsómetro es un monitor de la frecuencia cardíaca que consiste en una banda pectoral que detecta los impulsos del corazón y los envía –sin cables– a un reloj de muñeca, donde se visualizan.

- 1) ira-miedo-alegría
- 2) alegría-ira-miedo
- 3) miedo-alegría-ira

Si se fija, la ira se representa una vez la primera, otra vez después de la alegría y otra después del miedo: se representa en todas las posiciones posibles. Lo mismo ocurre con las otras dos "caras". Con esta estrategia de controlar el orden de presentación de los niveles de la variable independiente conseguimos que todos los valores de la TC que registramos cuando se pone una cara determinada se hayan producido en las mismas condiciones.

Realizamos la tarea de manera individual, en el mismo local, en el mismo momento del día, con el mismo instructor y teniendo cuidado de que los actores no vean el reloj con las pulsaciones. Agrupamos todos los valores medidos para cada una de las tres expresiones faciales y calculamos su media. Supongamos que se han obtenido los resultados siguientes (Figura 3.1.):

Figura 3.1. Respuesta fisiológica a las expresiones faciales



Con los resultados anteriores podemos decir que hemos conseguido repetir los resultados de Eckman et al (1983). Estos resultados nos dicen que nuestro cuerpo nos envía señales que están interpretadas de acuerdo con la emoción que está experimentando. Es algo así como si nuestra conciencia nos dijera: "fíjate qué triste estás que estás llorando". Comportarnos "como si estuviéramos bien" parece contribuir a sentirnos bien.

En la página web se dispone de una actividad complementaria (capítulo III, apartado 2.1).

Un profesor nuestro de psicopatología –José María Poveda– comentó en clase que él, a veces, a las mujeres depresivas, al acabar la primera sesión les decía que, como obligación, antes de la segunda sesión se pasaran por la

peluquería. Normalmente, además del pelo, habían hecho un extraordinario con la ropa; esto hacía que la sesión comenzase con elogios sobre su aspecto. Aparentar encontrarse mejor les ayudaba a sentirse mejor.

2.2. Ventajas de experimentar con los mismos sujetos: control y economía. Inconvenientes: aprendizaje, fatiga, motivación y práctica

Como recordará de los apartados anteriores, la mayor preocupación del investigador cuando desea formar los grupos de participantes para que respondan a cada una de las condiciones experimentales es que estos grupos estén lo más equilibrados posible. Para conseguirlo, primero controlaba el máximo de variables extrañas y después utilizaba el azar para generar los grupos y así equilibrar las diferencias individuales. Casi todas estas preocupaciones desaparecen si podemos hacer que el mismo grupo de participantes pase por todas las condiciones. Piense un momento en nuestro caso: si tenemos tres actores con un control excepcional de sus músculos faciales, esos mismos tres estarán equilibrados cuando les pidamos que muevan los músculos del miedo, los de la ira y los de la alegría. El posible efecto incrementador de su habilidad se mostrará por igual en todas las condiciones y no alterará las comparaciones de los resultados. Piense que el ideal es que si los niveles de la variable independiente producen el mismo efecto, entonces las respuestas de los grupos de sujetos a los distintos niveles sean iguales. Cuando los participantes que responden a las distintas condiciones son distintos, conseguir esto es trabajoso. Cuando son los mismos, casi podemos asegurar que se consigue. Por lo tanto, si la TC del mismo grupo de actores es distinta ante las diferentes solicitudes musculares, eso querrá decir que la única causa del cambio será el grupo muscular activado.

El control de las variables extrañas asociadas a las diferencias individuales es la ventaja mayor pero no la única. Conseguir participantes para las investigaciones no siempre es fácil (aunque es más fácil cuando sirven los estudiantes de la propia Facultad), y conseguir "buenos" participantes, menos. Por eso, con esta lógica intrasujetos, es mucho más fácil porque se necesitan menos; es más económico.

Utilizar los mismos participantes para todas las condiciones experimentales es la mejor forma de detectar la acción de la variable independiente.

Si es mejor para el control y más económico, ¿por qué no se hacen todos los diseños con los mismos participantes? Porque también existen grandes impedimentos. Empecemos por el primero. ¿Sabe conducir? Suponemos que sí (si no, elija otro contexto).

Sigamos suponiendo: lee un artículo en la prensa donde se refiere cómo un revolucionario método para aprender a conducir consigue reducir los suspensos en el primer examen a la mitad y, por si fuera poco, la formación adquirida es mejor. ¿Se animaría a probarlo? Su respuesta será que no. Ya no puede pasar por la segunda condición de una variable independiente llamada *métodos de aprender a conducir*. Ha aprendido. Ya no puede volver a la situación anterior (no podemos bañarnos dos veces en las aguas del mismo río). Repasemos nuestro experimento. ¿Aprenden nuestros actores de una condición a la siguiente? No, puesto que no hay una modificación sustancial en su organismo que les haga acometer la siguiente tarea más eficazmente.

Cuando se produzca aprendizaje en alguna de las condiciones experimentales será imposible aplicar la lógica intra.

Vayamos con el segundo impedimento. Nos disponemos a comprobar si las formas A y B de un nuevo test de cálculo numérico son igual de difíciles. La población objetivo son estudiantes de cuarto de Secundaria. Haciendo gala de nuestra mejor sonrisa aparecemos un viernes a las dos de la tarde, al finalizar la última clase del instituto. Les indicamos que sólo se tarda una hora y que... por supuesto, es voluntario quedarse. ¿Anticipa el resultado? ¿Qué falló? Seguramente la motivación. Será mejor estrategia separar la aplicación de las dos formas, hacerlo entre horas y quizá premiar con un bote de refresco o, mejor, con una entrada para ir al cine.

En la página web se dispone de una actividad complementaria (capítulo III, apartado 2.2).

Si vuelve unas páginas atrás encontrará que en el diseño de nuestro experimento de expresión facial nos preocupaba la duración total. En concreto decíamos: "en total la tarea durará veinte minutos y medio, teniendo en cuenta que diez minutos serán de descansos. Un tiempo que parece razonable para un actor". Esta preocupación era para evitar que la fatiga se acumulase al final del experimento y afectara a los resultados. Antes de hacer el experimento, habrá que preverlo. Si no tenemos información previa, lo mejor será hacer un experimento piloto (un experimento de prueba) para obtener información que permita mejorar el diseño definitivo.

Las tres amenazas anteriores (aprendizaje, motivación y fatiga) deberán ser analizadas antes de llevar a cabo el experimento con los mismos sujetos, ya que si se da alguna de las tres amenazas el experimento quedará invalidado.

Existen otras dos amenazas que se producen durante la realización del experimento y a las que debe estar atento el investigador. Estas dos amenazas son la *persistencia* y la *práctica*.

Entre una expresión facial y la siguiente dejábamos pasar medio minuto para que la TC volviera a la normalidad; si no lo hubiéramos hecho así, al empezar la expresión siguiente el incremento de la TC se sumaría al provocado por la condición anterior. Si la expresión de ira hace aumentar diez pulsaciones y –sin esperar a que bajen– a continuación solicitamos la expresión de alegría –que hace aumentar dos pulsaciones–, la medida que registraremos para la alegría será, falsamente, un aumento de doce. Se ha producido un efecto de persistencia.

Al comenzar una condición experimental tendremos que asegurarnos de que los efectos provocados por la condición aplicada anteriormente han desaparecido.

Esta precaución es especialmente pertinente cuando se investiga con sustancias que producen alteraciones en el organismo, como medicinas, drogas, dietas alimenticias, etc.

La persistencia a veces produce un cambio en la condición siguiente que no es posible reequilibrarlo con la ordenación inversa: en estos casos se habla de *persistencia diferencial*, que quiere decir que pasar del nivel uno al dos no es igual que pasar del nivel dos al uno. Observe la secuencia siguiente: aceite caliente-huevo-sal. Y ahora: sal-huevo-aceite caliente. Los efectos acumulados de la ordenación no son reversibles. Cuando esto ocurre, se considera que se ha alterado la naturaleza del experimento y que no es válido¹⁶.

La última amenaza que se produce siempre –en mayor o menor cuantía– es el efecto derivado de repetir varias veces tareas parecidas. No es lo mismo responder a una condición en primer lugar que hacerlo en último. Siempre habrá una ligera diferencia debida a la práctica que se lleva realizada. Recuerde cómo solucionábamos esta dificultad con las expresiones faciales: hacíamos que cada expresión se realizara en cada una de las posiciones posibles, 1ª, 2ª y 3ª.

Para contrarrestar el efecto de la práctica se controla el orden de presentación.

En el apartado siguiente presentamos varias de estas técnicas de control.

Antes de seguir, le sugerimos que se familiarice más con este tipo de experimentos realizando la actividad 3.2. de la página web.

16. Para ampliar conocimientos sobre diseños con control de la persistencia, puede consultar Jones y Kenward (1989).

En la página web se dispone de una actividad complementaria (capítulo III, apartado 2.2).

2.3. Tipos de diseños para controlar el efecto de la práctica: aleatorización simple; aleatorización en bloques; reequilibrado y cuadrado latino

Si ya ha realizado la actividad 3.2 se habrá dado cuenta de que para poder obtener información de cada uno de los niveles de procesamiento necesitamos hacer varias presentaciones de cada uno de los niveles de la variable independiente. ¿Qué pasaría si hacemos una sola pregunta en cada una de las condiciones? Lo más probable es que se recordaran las tres, no importa el nivel al que se hubieran procesado. Este tipo de investigación necesita muchas repeticiones y –afortunadamente– cada una consume muy poco tiempo. Cuando se dan estas circunstancias, para controlar el efecto de la práctica es suficiente con hacer una ordenación aleatoria simple del total de las presentaciones. Por ejemplo, en la actividad 3.2 tenemos tres niveles de la variable independiente; deseamos hacer seis repeticiones de cada nivel; en total tendremos dieciocho presentaciones. La apariencia final será el resultado de sortear el orden para esas dieciocho presentaciones. De este modo, la práctica se va repartiendo de manera aleatoria para cada una de las condiciones experimentales.

Si el número total de estímulos no es grande, o si no queremos confiar al azar el reparto de la práctica, se pueden ir haciendo ordenaciones aleatorias en bloques. Cada bloque contendrá el total de los niveles de la variable independiente. Por ejemplo, si la variable independiente tiene tres niveles (A, B, C) cada bloque tendrá tres presentaciones y éstas serán obligatoriamente los tres niveles de la variable. El orden de estas tres presentaciones para el primer bloque se hará mediante el azar: vg. B-C-A. Después prepararemos un segundo bloque siguiendo la misma lógica del primero; volvemos a sortear el orden para el segundo bloque: vg. C-B-A; y así sucesivamente hasta completar tantos bloques como el número de repeticiones que deseemos.

En las clases presenciales que impartimos en la UAM los estudiantes están divididos en dos subgrupos para hacer las prácticas. Al acabar un tema, el subgrupo "uno" debe tener preparado su trabajo para el "día de prácticas", que es uno a la semana. Hasta la semana siguiente no lo presentará el subgrupo "dos". Para evitar que se enmascare un efecto extraño al hecho de ser los pri-

meros en exponer al final de cada tema (menos tiempo, no conocer las correcciones hechas a los demás, etc.), al cambiar de tema el orden de las prácticas se reequilibra: primero el subgrupo "dos", y a la semana siguiente el "uno". La comprobación de si conseguimos controlarlo efectivamente la hicimos contrastando, al final de curso, que los dos subgrupos sacaron notas similares.

Cuando la variable independiente tiene dos niveles y las tareas que han de realizar los participantes son largas, la técnica adecuada es el *reequilibrado*¹⁷.

En el diseño con reequilibrado se busca que la descompensación (desequilibrio) que produce una ordenación determinada de las condiciones experimentales sea compensada (reequilibrada) con la ordenación simétricamente inversa. Cuando son dos niveles (A, B) esto se consigue fácilmente: la ordenación A-B es reequilibrada con la ordenación B-A. En este diseño, al saberse cómo se respondió a A cuando se presentó en primer lugar y cuando se presentó en segundo lugar, –lo mismo para B– puede calcularse el efecto de responder "en primer lugar" y responder "en segundo lugar". Conocer matemáticamente este efecto ayuda a dejar más aislado y más claro el efecto debido a la variable independiente. Un ejemplo de control por reequilibrado lo tiene en el caso 3.2 y la actividad 3.3. complementaria en la página web.

En la página web se dispone de una actividad complementaria (capítulo III, apartado 2.3).

Caso 3.2

¿Un poco más de sufrimiento leve hace que se sufra menos?

Por un momento trate de ser lo más realista que le sea posible. Su médico de cabecera ha detectado algo raro en su corazón. Para probarlo le pide que le acompañe a una pista de atletismo donde deberá correr tan sólo doce minutos. Eso sí, deberá hacerlo a la mayor velocidad posible. No vale pararse. Así que empieza a correr fuerte, pero pensando que quedan más de diez minutos. Cuando quedan cinco minutos, y ya resopla, sudando como en una tarde de julio, su médico le dice que se esfuerce un poco más, que seguro que puede. Lo intenta, pero le empiezan a doler las piernas y siente algunas punzadas debajo de las costillas. Cuando queda un solo minuto, y lleva ya tiempo respirando con la boca lo más abierta posible, escucha una voz interior que le dice que no puede más y que se pare. Es ahora cuando el médico le jalea y le dice que si no hace un *sprint* en este minuto, todo habrá sido inútil. En la última recta, hace tiempo que ha dejado de sentir las piernas, le duele el pecho y siente que se ahoga, cree que se va a caer de un momento a otro y tiene la sensación de que sólo se mantiene en pie por los constantes gritos del

17. Del inglés *balance*, equilibrio, *counterbalance*, reequilibrio.

médico que le dice "un poco más", "ya acabas". Justo al llegar a la meta, el "médico-torturador" le comenta que se descansa mejor si se hacen otras tres vueltas –de cuatrocientos metros– al trote, pero que si lo prefiere, puede dejarlo ¹⁸.

¿Qué cree que haría usted?

La mayoría de las personas pararían. Sin embargo, los corredores, profesionales o aficionados, cuando entrenan, siguen corriendo. Aparte de las razones fisiológicas, existen razones psicológicas de experiencia del dolor, puestas a prueba experimentalmente por Kahneman, Frederickson, Schreiber y Redelmeir (1993). Estos autores comprobaron el efecto que tenía sobre la experiencia del dolor un periodo añadido de dolor con una intensidad menor.

El experimento tenía como variable independiente la duración de la experiencia dolorosa, con dos niveles: en el primer nivel había que tener una mano sumergida en agua a catorce grados durante un minuto¹⁹; en el segundo nivel, después del minuto en agua a catorce grados, seguía un periodo de medio minuto en agua a quince grados.

Los autores quisieron controlar al máximo las variables extrañas asociadas a las diferencias individuales, por lo que diseñaron el experimento con los mismos participantes. Primero pasarían por la experiencia corta y después por la larga. Sin embargo, esta lógica tenía dos amenazas. La primera era que podía producirse un efecto de persistencia, haciendo que en la segunda aplicación la sensibilidad dolorosa se viera afectada por los efectos producidos en la primera; para solucionarlo determinaron el tiempo necesario entre condiciones para que la sensibilidad retornara a la normalidad: siete minutos. La segunda amenaza era que podía producirse un efecto de adaptación o de práctica que hiciera que la segunda condición resultara distorsionada; para solucionarlo, controlaron el orden de manera reequilibrada: la mitad de los participantes hizo el experimento en la ordenación corto-largo y la otra mitad en la ordenación largo-corto.

Cuando se preguntó a los participantes cuál de las dos experiencias preferirían repetir, el 69% dijo que la experiencia larga.

Puede pensarse como explicación alternativa que 14 grados frente a 15 grados es límite del dolor, siendo la primera dolorosa y la segunda no. Para poder descartar esta explicación, los autores evaluaron de forma independiente la sensación producida por el agua a quince grados, habiendo sido calificada como dolorosa.

Ahora, por favor, vuelva a leer en el caso con el que comenzábamos este apartado (apartado 2.1: "Presentación de un caso"), la parte relativa al orden en que los actores debían configurar cada una de las expresiones faciales.

18. Al finalizar, los músculos se encuentran en deuda de oxígeno por el esfuerzo anterior y con gran cantidad de lactato. Si se permite que el corazón siga bombeando a ritmo moderado un poco más y moviendo los músculos, se consigue mejor el equilibrio y se elimina más lactato.

19. Las temperaturas del agua no producen la misma sensación que las del aire. A catorce grados, con un traje de neopreno de 9 mm se pasa un frío de mil demonios. Haga, si no, la prueba.

En la página web se dispone de una animación complementaria (capítulo III, apartado 2.3).

Cuando el investigador tiene una variable independiente con más de dos niveles y desea que se produzcan el mínimo de repeticiones de cada nivel, entonces la estrategia que tiene que seguir es que cada nivel se presente una vez en cada posición posible. Esta estrategia recibe el nombre de *cuadrado latino*. Ésta es la estrategia que seguimos en la presentación de las emociones:

Tabla 3.1.

	Posiciones		
	1.ª	2.ª	3.ª
Ordenación 1	Ira	Miedo	Alegría
Ordenación 2	Alegría	Ira	Miedo
Ordenación 3	Miedo	Alegría	Ira

Hay otra forma de hacer la presentación en cuadrado latino con estas tres emociones. Si se anima a intentarlo puede verlo con la actividad 3.5.

Una de las ventajas adicionales de este diseño en cuadrado latino es que en el análisis de datos posterior puede calcularse la cuantía del efecto debido al orden.

En una investigación en la que cada una de las condiciones consume poco tiempo, como la nuestra, todos los participantes pueden pasar por todas las ordenaciones planeadas. Se dice entonces que el diseño es *completo*. Cuando no es posible, se divide –al azar– a los participantes en subgrupos, de forma que cada subgrupo responda a los niveles de la variable en una de las ordenaciones planeadas. En nuestro caso, si fuera muy largo, un tercio de los actores haría las tres expresiones en la ordenación "ira-miedo-alegría", otro tercio en la ordenación "alegría-ira-miedo" y el último tercio en la ordenación "miedo-alegría-ira". Se dice entonces que el diseño es *incompleto*.

Fíjese en que, aunque todos los participantes no responden a las condiciones en todas las posiciones posibles, sí está asegurado que todos los participantes responden a todas las condiciones; cada ordenación contiene, por definición, todos los niveles (de no ser así, el diseño no podría llamarse *con los mismos sujetos*). Esta misma consideración se hace cuando el diseño es reequilibrado, con dos niveles. Puesto que hay dos ordenaciones posibles, los participantes pueden hacer las dos o, divididos en dos subgrupos, realizar cada subgrupo una.

3. Experimentos con un solo sujeto

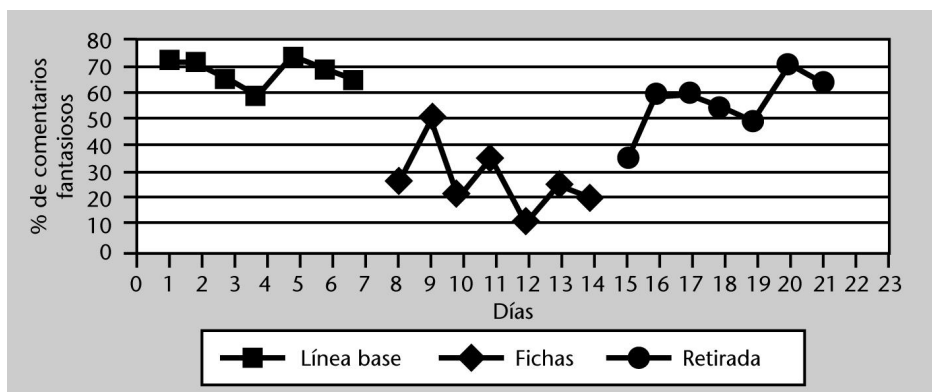
3.1. Presentación de un caso. El control; la intervención y la retirada. Tipos de diseños con un solo sujeto

En el ámbito de la Psicología Clínica surge la necesidad de conocer la eficacia de una determinada terapia pero no se dispone de un grupo de pacientes similares con los que plantear una estrategia experimental con distintos o los mismos sujetos. Sólo hay un paciente con el que experimentar... siempre y cuando se consiga el objetivo primordial: ayudarle a cambiar su conducta. En estos entornos se hace necesario un plan de investigación experimental que tenga un solo participante. Empecemos viendo un caso de lo que estamos diciendo.

Uno de los programas usados en modificación de conducta consiste en premiar –reforzar positivamente– una conducta mediante la entrega de fichas, las cuales tienen un determinado valor acordado. La tesis que subyace es que la conducta reforzada positivamente tenderá a repetirse. Las fichas tienen atractivo para el paciente porque, por ejemplo, pueden cambiarse por productos de la cafetería del hospital. Adicionalmente, puede reforzarse negativamente una conducta que se desea eliminar.

Wincze, Leitenberg y Agras (1972) reforzaron, mediante fichas, "no hablar de cosas fantasiosas", en un enfermo crónico de esquizofrenia paranoide. El periodo de aplicación de la terapia de conducta fue durante las entrevistas diarias que el terapeuta tuvo a lo largo de veintidós días con el paciente. Durante los siete primeros días se limitó a anotar los comentarios fantasiosos, sin hacer nada. En los siete días siguientes aplicó el programa de refuerzo mediante fichas. Comprobado que el porcentaje de comentarios fantasiosos había disminuido, en el periodo de los siete días siguientes volvió a una situación similar a la de la línea base. Como podemos observar en la figura 3.2, cuando retiran el sistema de refuerzos al enfermo, éste vuelve a hacer más comentarios fantasiosos.

Figura 3.2. Efecto de refuerzo mediante fichas sobre el habla fantasiosa



Una vez que comprobaron que el sistema de refuerzos mediante fichas era la causa de la disminución de la conducta indeseada, a partir del día 22, volvieron a la terapia para reducir los comentarios fantasiosos.

En este caso pueden apreciarse los elementos característicos de los experimentos con un solo sujeto: en primer lugar el establecimiento de la línea base, después la fase de intervención y finalmente la fase de retirada.

En el caso presentado, cuando los investigadores anotan la cantidad de pensamientos fantasiosos que tiene el paciente, lo que están haciendo es establecer un control de su conducta. Veamos por qué. Si tuviéramos un grupo de enfermos, podríamos dividir la mitad al azar para ser tratados y la otra mitad para que sirviera de control. En el grupo de control habrá algunos pacientes con más pensamientos fantasiosos y otros con menos. Al investigador, esto no le preocupa mucho, porque lo que comparará después será la media del grupo. Es algo así como establecer el enfermo prototípico.

Cuando tenemos un solo paciente nos arriesgamos a que la medida de control que tomemos no represente bien su estado. Por ejemplo, si sólo tomamos la medida del día 5 nos encontramos un 75% de ideas fantasiosas, mientras que si lo hubiéramos hecho el día 7 habríamos registrado un 65%. Un investigador alocado podría pensar que se está produciendo una recuperación natural y que no hace falta intervenir. Sin embargo, cuando tenemos todos los registros de la semana –la línea base– nos damos cuenta de que esa disminución forma parte de la variación natural de la conducta.

El investigador necesita saber cómo es la conducta durante un periodo de tiempo antes de ser tratada, para poder interpretar después los posibles cambios de ésta. Si observa ahora los registros durante la fase de intervención, observará que se produce todavía una variación más grande. Sin embargo,

viendo todas las puntuaciones, el terapeuta puede deducir que hay una tendencia media a la disminución del porcentaje de pensamientos fantasiosos: se pasa del 60% en el mejor de los días sin terapia a un porcentaje del 20% el último día de la semana con terapia. Viendo el conjunto de la fase de intervención, el experimentador interpreta que ese 20% no se debe a una variación natural, ni a una puntuación extrema, sino que es la evolución del paciente sometido a la terapia de modificación de conducta mediante fichas.

Antes de seguir queremos hacerle notar cómo, aunque los nombres cambian, la fase de línea base coincide conceptualmente con una condición de control y la fase de intervención con la condición experimental.

En la fase de línea base hay un nivel nulo de la variable independiente, y en la segunda, un nivel "tratamiento X", cuyo efecto queremos comparar con el control. Por lo tanto, estamos dentro de la lógica experimental, aunque tengamos un solo participante.

En la práctica clínica, cuando se observa que un paciente mejora ostensiblemente, la terapia continúa hasta alcanzar un criterio óptimo fijado con anterioridad. Cuando a la práctica clínica se une la exigencia investigadora es necesario añadir algo más. Es necesario poder contestar a otros colegas preguntas como las siguientes:

- a) ¿Hubo alguna circunstancia personal, en paralelo, que pudiera explicar la mejoría del paciente?
- b) ¿Se dio una especial empatía terapeuta-enfermo que explicara su recuperación?
- c) ¿No coincidiría la mejoría con el final de otoño, después de la época mala para las enfermedades mentales?
- d) Por supuesto, ¿no habría un cambio en la medicación por esas fechas?

En la página web se dispone de una actividad complementaria (capítulo III, apartado 3.1).

Si se fija atentamente, todas estas amenazas se enmascaran con la aplicación de la terapia y rivalizan con ella en la explicación de la mejoría. Algunas de ellas podrán ser rechazadas recopilando la información correspondiente. Pero otras no. Para poder rechazarlas todas es necesario ampliar la estrategia del experimento retirando la intervención. Si aparece un deterioro claro, coincidiendo en el tiempo con la retirada de la terapia, entonces podremos estar más seguros de adscribir el cambio de la conducta a la intervención del terapeuta. Si observa los datos del paciente presentado, hay un deterioro claro que se aproxima a valores de la línea base inicial.

Suponemos que no se imagina que una vez que se comprobó que la terapia de fichas ayudaba al paciente esquizofrénico la investigación acabó en fase de retirada. Una vez que los investigadores se han cerciorado de su efectividad vuelven al tratamiento hasta alcanzar su objetivo o hasta que la terapia alcanza su límite natural.

En el entorno de los diseños experimentales con un solo participante se ha desarrollado una terminología propia para hablar de forma simplificada de los diseños. Las fases sin tratamiento se designan con la letra A (aunque sea la primera de línea base), y las fases con tratamiento, con la B.

Chudy, Jones y Dickson (1983) utilizaron un diseño AB para tratar una conducta fóbica. Se trataba de un niño de siete años que había sido agredido por un pastor alemán. Como consecuencia desarrolló una conducta fóbica hacia todos los perros, miedo a salir solo de casa y disminución de sus contactos con otros niños. Tras un periodo de veintiún días de línea base, se le entrenó en una desensibilización sistemática –modificada con imaginación emotiva, juego libre y modelado–. Al finalizar la terapia, se observó que era capaz de acariciar un perro y que había aumentado el número de veces que salía solo de casa, así como el número de veces que se relacionaba con otros niños.

Este diseño AB seguramente no les pareció mal a los padres del niño, ya que el objetivo de su recuperación se cumplió. Sin embargo, a los investigadores nos parece que presenta amenazas a su validez. El plan recomendado tendría que ser un diseño ABAB. Con esta lógica de las letras, no es difícil entender los diseños que los investigadores de este campo han ido extendiendo.

En el caso 3.3 puede ver extensiones de la lógica ABAB.

Caso 3.3

Extensiones de la estructura ABAB de los diseños con un solo sujeto

1) **Incorporación de varios niveles de la variable independiente: A-B-B'-B''** (Hartmann y Hall, 1976).

B, B' y B'' designan tres niveles de intensidad creciente en una variable de intervención. Una vez comprobado que se ha producido la primera mejora entre A y B, se utiliza B como línea base para contrastar un supuesto cambio favorable hacia B'. Con la misma lógica se compara B' con respecto a B''.

2) **Incorporación de dos tratamientos alternativos: A-B-A-C-A** (Wincze, Leitenberg y Agras, 1972).

B y C representan los dos tratamientos alternativos cuya eficacia comparada se desea contrastar. Con A-B-A se pone a prueba la efectividad del primero. Con A-C-A se pone a prueba la acción del segundo. Observe que la retirada tras la primera intervención deberá mantenerse lo suficiente para que no se produzca un efecto de persistencia que pudiera enmascarar una mayor eficacia de C.

3) **Incorporación de dos tratamientos superpuestos: A-B-C-A-BC-C**

(Bernard, Kratochwill y Keefauver, 1983).

Con esta lógica se busca comparar si la acción simultánea de BC supera las mejoras de B y C por separado.

Si desea profundizar más sobre diseños con un solo sujeto, debe acudir a textos específicos de esta área, como los de Barlow y Hersen (1984) y Kazdin (1982).

Las claves del control en los diseños con un solo sujeto son las series de medidas y la retirada del tratamiento.

4. Experimentos con más de una variable independiente: diseños factoriales

4.1. Presentación de un caso

En el momento de preparar este apartado nos hallamos sentados frente a la bahía de San José (Almería). Ha caído la noche y las luces del pueblo cercan el negro del mar fundido con el negro de un cielo sin luna. En este "contexto" hemos recordado un experimento de Godden y Baddeley (1975) sobre si se recuerdan mejor las cosas cuando se está en el mismo "contexto" en el que se estudiaron. Los contextos en que investigaron fueron tan sugerentes como "la playa" y "debajo del agua". Con estas ideas vamos a diseñar un experimento factorial que esperamos que le resulte atractivo. Pero vayamos despacio desde el principio. Nuestro objetivo es planear dos experimentos de una variable independiente, como los que hemos presentado hasta ahora. La novedad va a consistir en hacerlos de forma simultánea.

En el primero de ellos se va a poner a prueba la hipótesis siguiente: "si da lo mismo estudiar debajo del agua que en la playa, entonces el recuerdo posterior de lo estudiado será igual". A esta variable independiente la llamamos *tipo de "contexto"*.

En el segundo experimento simultáneo se va a poner a prueba la hipótesis siguiente: "si el paso del tiempo debilita la huella en la memoria, una hora después de haber memorizado se recordará menos que inmediatamente después". A esta segunda variable independiente la llamamos *demora*.

La variable dependiente será el recuerdo medido como el porcentaje de palabras reconocidas correctamente. Aunque en el texto hablemos de recuerdo para una mayor claridad, tendremos que entender siempre que se trata de re-

conocimiento; es decir, un par de minutos después (o una hora después) del tiempo dejado para memorizar, se les presentará a los participantes una lista con palabras en las que unas corresponderán a las memorizadas y otras serán de relleno. Lo que tendrán que hacer todos los grupos será marcar con un lápiz las palabras que coincidan con las presentadas inicialmente.²⁰

¿Dónde le gustaría preparar el próximo examen? ¿Qué le parece repasar el glosario de esta materia mientras cientos de fosforescentes castañuelas²¹ azules pasan a su alrededor? ¿Preferiría hacerlo al borde del agua mientras toma el sol? Veamos cómo distribuiríamos a los participantes para nuestros dos experimentos simultáneos: supongamos que hemos decidido aplicar la lógica del diseño de grupos aleatorios, con distintos sujetos; necesitaremos voluntarios para encuadrarlos en cada una de las cuatro condiciones posibles:

- 1) Contexto bajo el agua-recuerdo inmediato.
- 2) Contexto bajo el agua-recuerdo demorado.
- 3) Contexto en la playa-recuerdo inmediato.
- 4) Contexto en la playa-recuerdo demorado.

Si deseamos tener diez personas por grupo necesitaremos un total de cuarenta. Repasando la lógica de los experimentos recordamos que todos ellos deberán estar igualados en las variables extrañas que pudieran afectar a la variable dependiente. ¿Cree que cualquiera valdría para participar en la condición "bajo el agua"? Para algunas personas es una situación muy estresante. Por lo tanto, para controlar esta variable, podríamos elegir a todos los participantes entre personas que bucean habitualmente. Suponga que tenemos los ciento cincuenta miembros de un club de buceo de L'Estartit (Girona) dispuestos a participar. Seguimos controlando variables: dejamos fuera a los que tengan algún problema de salud actual; no discriminamos a nadie por edad; les pasamos un test de memoria verbal y nos quedamos con los que tienen valores en torno al 90% de la media poblacional y, por último, igualamos el número de varones y mujeres. Suponga que este número resulta ser cuarenta y cinco. Para dejarlo en cuarenta, sorteamos qué cinco se quedan fuera.

Ya tenemos a nuestros cuarenta voluntarios dispuestos a memorizar, bien en la playa, bien debajo del agua. A continuación tenemos que hacer la asignación aleatoria de los cuarenta buzos a las cuatro condiciones. En la Tabla 3.2 presentamos un esquema con la distribución de los participantes por condiciones.²²

20. Los lápices escriben debajo del agua.

21. Castañuelas: Especie muy corriente en el Mediterráneo, que cuando es pequeña tiene un característico azul brillante.

Tabla 3.2.

		Variable independiente: Demora		
		Inmediato	Una hora	Resultados Contexto
Variable independiente: Contexto	Bajo el agua	Diez participantes estudiarán bajo el agua y se les preguntará de inmediato.	Diez participantes estudiarán bajo el agua y se les preguntará una hora después.	% de palabras que reconocerán los que estudien bajo el agua.
	En la playa	Diez participantes estudiarán en la playa y se les preguntará de inmediato.	Diez participantes estudiarán en la playa y se les preguntará una hora después.	% de palabras que reconocerán los que estudien en la playa
	Resultados Demora	% de palabras que reconocerán los que sean preguntados de inmediato.	% de palabras que reconocerán los que sean preguntados una hora después.	

Preparamos los materiales necesarios para la realización de la tarea, que consistirán, para los dos grupos submarinos, en una lista de treinta palabras escritas en pizarritas con suficiente tamaño y claridad para leerse debajo del agua. A cada buzo se le proporcionará una pizarrita que tendrá guardada en uno de los bolsillos del chaleco hidrostático. Por supuesto, comprobaremos que las botellas tienen la suficiente presión de aire –no queremos perder participantes de forma no aleatoria.

Para los dos grupos que memorizarán en la playa, dispondremos de la misma lista de palabras, pero esta vez escrita en cartones individuales. Para estos segundos grupos se dispondrá de la correspondiente crema para el sol.

También hay que prever una tarea de distracción para los grupos que realizarán la prueba de reconocimiento con una demora de una hora. Para los que estén en el agua, nada mejor que una placentera inmersión, sin bajar mucho –máximo quince metros–, para evitar que las posibles molestias de la descompresión puedan afectarlos. Para los que se queden en la playa, un partido de

22. Recuerde: la lista alfabética, la numeración, el número random de la calculadora y el reparto. Ya sabe que en la actividad 3.1 se muestra en acción cómo hacerlo.

balonvolea puede cumplir el mismo objetivo. Observe que los dos grupos de demora deben hacer una actividad parecida, no podemos hacer que unos estén esquivando tiburones y los otros leyendo a Marsé.

Todos los grupos tratarán de memorizar el mayor número de palabras durante un tiempo de dos minutos. Dos minutos después se medirá el reconocimiento en los dos grupos sin demora. Sobre una lista de noventa palabras deberán marcar las treinta que reconozcan de la primera lista. Al cabo de una hora se procederá de la misma manera con los otros dos grupos.

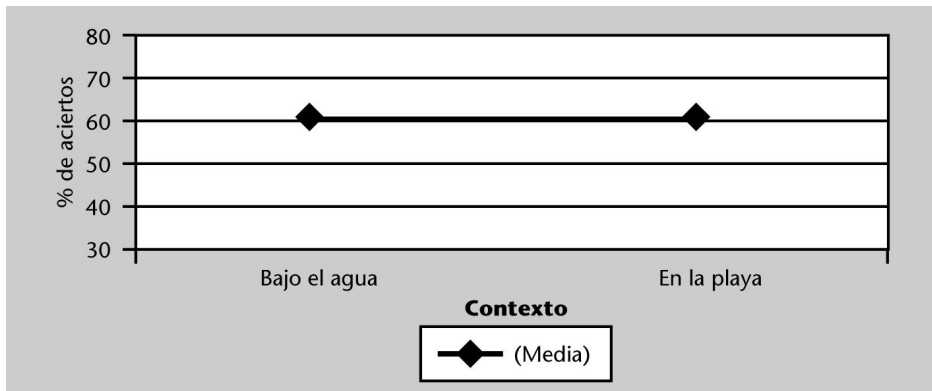
Para situar los resultados utilizamos la planilla que aparece en la Tabla 3.2, donde anotamos el porcentaje de palabras correctamente reconocidas, de promedio, por cada uno de los cuatro grupos. Hemos simulado unos resultados que aparecen en la Tabla 3.3.

Tabla 3.3.

		Variable independiente: Demora		
		Inmediato	Una hora	Resultados Contexto
Variable independiente: Contexto	Bajo el agua			60%
	En la playa			60%
	Resultados Demora	70%	50%	

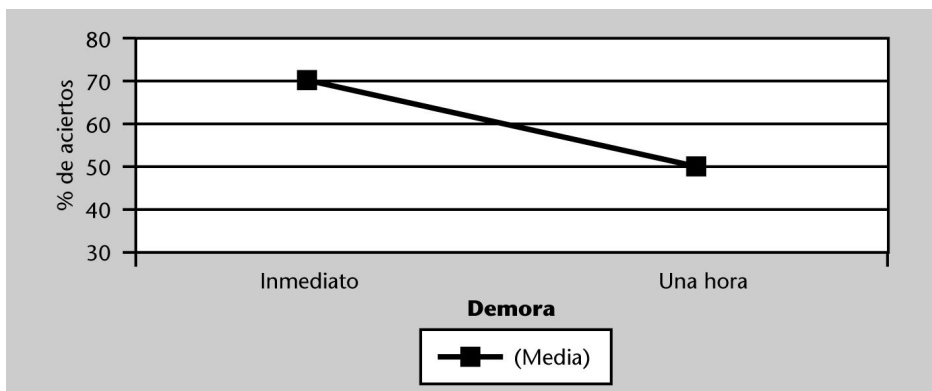
Vayamos ahora a comentar los resultados de cada experimento. Empecemos con el "contexto" en que se hace la tarea. Los resultados en la columna derecha de la Tabla 3.3 nos dicen que da igual dónde se estudie: el grado de recuerdo (reconocimiento) de palabras es exactamente igual en la playa que debajo del agua.

Figura 3.3. ¿Cómo se rinde más? Desde el punto de vista del contexto de estudio



Para la variable siguiente, "demora", los resultados nos dicen que, claramente, se recuerdan (reconocen) menos palabras cuando ha pasado una hora (datos de la última fila de la tabla 3.3).

Figura 3.4. ¿Cómo se rinde más? Desde el punto de vista del paso del tiempo



Por lo tanto, si vemos este experimento factorial como dos experimentos de una variable independiente, tenemos que concluir lo siguiente:

- Primero, el "contexto" donde se memoriza y se recuerda no tiene incidencia en el reconocimiento posterior.
- Segundo, la "demora", como el tiempo transcurrido entre la memorización y el recuerdo, sí es relevante. Son mejores los resultados inmediatos que los obtenidos una hora después.

En los párrafos que siguen vamos a mostrarle lo siguiente: la mayor ventaja de un experimento factorial de dos variables es que obtenemos más información que la suma de la información de los dos experimentos de una variable implícitos.

¿De dónde se obtiene esta información extra? De las celdas que combinan la información de las dos variables independientes; de saber qué pasó, por ejemplo, con el subgrupo que estudió en la playa y al que, una hora después, se pidió que señalaran las palabras que reconocieran de la lista original. Para estudiar esta información necesitamos obtener las medias de los porcentajes de palabras reconocidas en cada una de las cuatro condiciones que se generaron. Esta información aparece ahora en la Tabla 3.4.

Tabla 3.4.

		Variable independiente: Demora		
		Inmediato	Una hora	Resultados Contexto
Variable independiente: Contexto	Bajo el agua	60%	60%	60%
	En la playa	80%	40%	60%
	Resultados Demora	70%	50%	

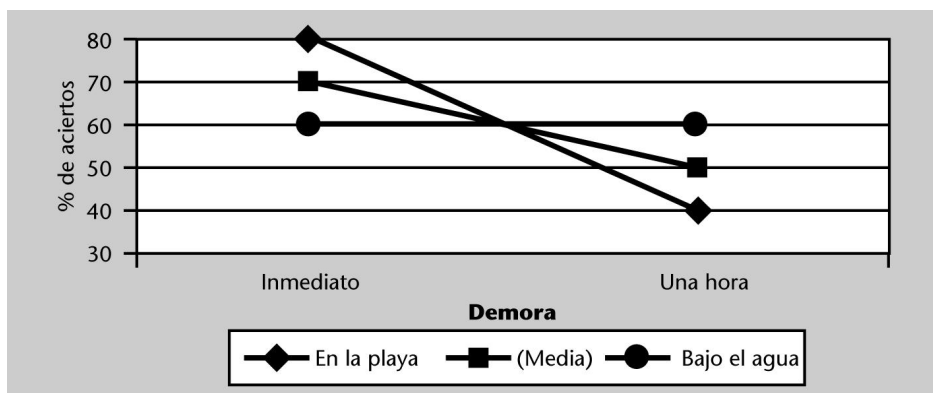
Para proceder al estudio de esta información debemos hacer dos representaciones gráficas, una para cada variable independiente²³

En la página web dispone de una herramienta que le enseña paso a paso cómo realizar las gráficas de los resultados factoriales. Para que el aprendizaje se complete, debe acompañar las explicaciones con la realización de la gráfica con lápiz y en papel cuadriculado. Cuando lo desee, puede activar la actividad 3.3.

En la página web se dispone de una actividad complementaria (capítulo III, apartado 4.1).

Figura 3.5. ¿Cómo se rinde más? Desde el punto de vista de "la demora", teniendo en cuenta el lugar en el que se ha estudiado

23. Para un estudio de todos los casos posibles en un diseño 2 x 2, puede consultar León y Montero (2001).



Observando ahora los resultados de la variable "demora", combinada con "contexto" donde se memoriza, podemos concluir dos cosas:

- La memoria –en términos de reconocimiento de palabras– se deteriora bastante (de 80% a 40%) al cabo de una hora, cuando se hace en el "contexto" de la playa (línea con rombos).
- La memoria –en términos de reconocimiento de palabras– no se ve afectada (de 60% a 60%) al cabo de una hora, cuando se hace en un "contexto" bajo el agua (línea con círculos).

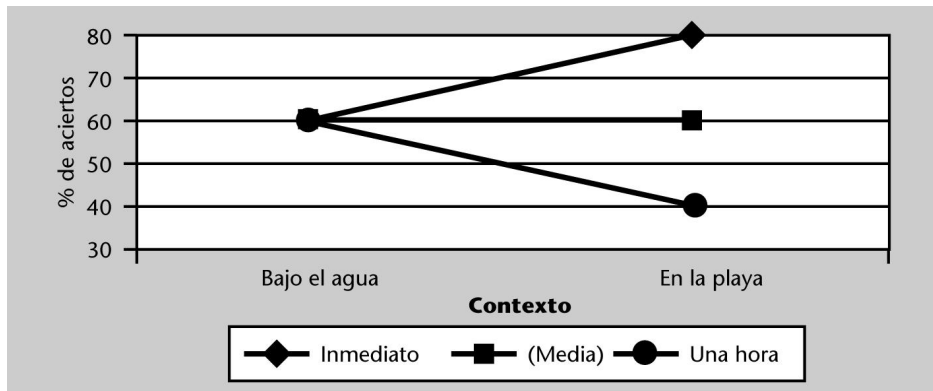
Como consecuencia de lo anterior, el promedio no refleja correctamente lo que ocurrió en el experimento para la variable "demora". El promedio dice que la memoria de palabras reconocidas al cabo de una hora se deteriora, para el conjunto de los contextos; cuando en realidad sólo se produce descenso para los que estaban en la playa; para los que lo hacen todo debajo del agua no hay deterioro de la memoria en el plazo de una hora.

Cuando el promedio no representa adecuadamente a la variable independiente, no se debe interpretar.

Recuerde lo de que "si yo me como dos salmones y un amigo mío ninguno, la media dice que cada uno nos hemos comido un salmón de promedio", lo cual es matemáticamente correcto, pero engañoso –sobre todo para mi amigo– y, por lo tanto, no nos sirve para nada. Esto quiere decir que cuando las dos líneas tengan pendiente de signo diferente (una hacia arriba y otra hacia abajo), el promedio no es un buen resumen de lo que ocurre con ambas variables independientes y, por lo tanto, no se debe interpretar.

Pasemos ahora a representar en la Figura 3.6 los datos desde el punto de vista de la otra variable: el "contexto". La situamos en el eje de las abscisas y representamos los valores de las cuatro condiciones, siguiendo todos los pasos explicados en la anterior variable independiente.

Figura 3.6. ¿Cómo se rinde más? Desde el punto de vista del contexto en el que se estudia y teniendo en cuenta el tiempo de demora



Observando ahora los resultados de la variable "contexto", combinada con "demora", concluimos dos cosas:

- Si se pregunta inmediatamente después de memorizar, se reconocen más palabras (80% frente a 60%) cuando el "contexto" en que se realiza el reconocimiento es la playa (línea con rombos).
- Si se pregunta una hora después de memorizar, se reconocen más palabras (60% frente a 40%) cuando el "contexto" en que se realiza el reconocimiento es bajo el agua (línea con círculos).

En la página web se dispone de una actividad complementaria (capítulo III, apartado 4.1).

Como consecuencia de lo anterior, el promedio de resultados del "contexto" no refleja correctamente lo que ocurrió en el experimento. El promedio indica que se producía el mismo grado de reconocimiento en la playa que debajo del agua. Se estaban compensando resultados opuestos: cuando se pregunta inmediatamente, se recuerda más en la playa, y cuando pasa una hora se recuerda más si se está debajo del agua.

Como comentario final podemos decir que, cuando interpretábamos los resultados correspondientes a dos experimentos de una VI –utilizando los promedios–, nos veíamos obligados a concluir que "en el plazo de una hora se aprecia un deterioro en la memoria" y que "se estudiara y recordara en el contexto que fuera, se recordaba siempre igual". Estas conclusiones no eran correctas –como acaba de ver– porque el promedio contradice el sentido de uno o los dos resultados parciales dentro de cada una de las variables independientes. Cuando se analizan los resultados desde la perspectiva factorial, podemos ver en detalle cómo han sido los resultados de las combinaciones de las variables independientes –además de los resultados promedios–. Por lo tanto:

Cuando estemos ante un experimento factorial de dos (o más) variables independientes, habrá que comprobar si la acción de una variable permanece igual cuando se combina con los niveles de la otra variable independiente.

Cuando una variable no cambia su efecto al combinarse con otra, diremos que *no se ha producido interacción*. Cuando esto ocurre las líneas son paralelas. Entonces podemos interpretar correctamente la acción de la variable independiente con sus resultados promedios.

Si la acción de una variable independiente sobre la dependiente es modificada al combinarse con los niveles de la otra variable independiente, diremos que *se ha producido una interacción*; entonces las líneas no son paralelas.

Cuando se produce interacción, no podemos estar seguros de que el promedio represente bien los resultados parciales. Habrá que comprobarlo. En el caso con que hemos presentado este apartado ninguna de las dos variables independientes podía interpretarse apoyándose en sus valores promedios. Esto ocurría porque el promedio contradecía algún resultado o los parciales (tenían las dos líneas pendientes de diferente signo).

Cuando la representación de los resultados son líneas paralelas, no existe interacción y puede interpretarse el promedio. Cuando no son paralelas existe interacción y hay que comprobar si se puede interpretar el promedio: cuando sus pendientes tienen el mismo signo, sí se puede interpretar.

Por último, si estuviéramos redactando las conclusiones de esta investigación simulada, deberíamos decir que se observa que en la playa se memoriza mejor si la memoria se va a poner a prueba de una forma inmediata. Esto podría explicarse porque es más cómodo, comporta menos estrés, y más familiaridad con los contextos de memorización habituales; descriptores que son las diferencias entre el contexto de la playa y el submarino.

Sin embargo, al cabo de una hora, en la playa se produce mayor olvido; recordar en el mismo contexto en que se estudió no ayuda mucho a los que memorizaron en la arena; para los que estudiaron bajo el agua, recordar en un contexto tan especial como en el que memorizaron hace que una hora des-

pués se recuerde tanto como de forma inmediata. Es como si el entorno tan diferenciado y las condiciones fisiológicas especiales se asociaran a lo que memorizaban y después sirvieran de coadyuvantes en el recuerdo.

Fijese que extraer conclusiones de modo parcial dentro de una variable independiente no hace perder fuerza a la investigación. Simplemente se concreta qué pasa bajo qué condiciones.

Todo esto lo expondremos de forma general y paso a paso en el apartado siguiente.

4.2. Los elementos de un diseño factorial: la representación gráfica, los efectos simples, la interacción y los efectos principales

Un *diseño factorial* es un diseño en el que el investigador estudia simultáneamente dos (o más) variables independientes. En el caso presentado se estudiaban las variables independientes: "contexto" de memorización y tiempo de "demora" en el recuerdo.

En un diseño factorial completo se combinan todos los niveles de una variable con todos los niveles de la otra o las otras. En nuestro caso, los dos contextos de memorización se combinaban con los dos tiempos de demora. Esto se representa mediante la multiplicación del número de niveles de las variables independientes. En nuestro caso, 2×2 .

Para comprender mejor los resultados de una investigación factorial de dos variables se deben representar en una gráfica. En realidad, en dos, una para cada variable independiente.

Se empieza realizando una gráfica para cualquiera de las dos variables independientes. La variable independiente elegida la consideramos como "central" a efectos de representación. El carácter de central lo da el hecho de figurar en el eje de las abscisas. La otra variable independiente actúa como "condicionante", ya que la central actúa bajo sus condiciones.

En nuestro caso, si empezamos con la representación de la variable "demora", considerada como central, nos referimos a la Figura 3.5. Después se realiza otra gráfica para la segunda variable independiente, considerada ahora como central. En nuestro caso, la gráfica para la variable "contexto", considerada como central, es la Figura 3.6.

Una vez que tenemos las gráficas para cada una de las variables independientes, empezamos a estudiar cómo se comporta la variable cuando se combina con la otra.

A estos resultados parciales de la variable independiente, agrupados bajo las condiciones de la otra variable independiente, se les da el nombre de *efectos simples*.

En nuestro caso, analizamos la variable "demora" combinada con los niveles del "contexto" (Figura 3.5).

Veamos un efecto simple de la demora: cuando se estudió bajo el agua, que está representado por la línea con círculos. El otro efecto simple de la demora: cuando se estudió en la playa, que está representado por la línea con rombos. (Para esto nos sirven las líneas que unen resultados bajo la misma condicionante).

Se les llama *efectos simples* porque es como si dividiéramos el experimento de una variable en dos experimentos simples.

En el primer experimento simple se investiga el efecto de la "demora", condicionado a que todos los participantes estudien bajo el agua.

En el otro experimento simple se investiga el efecto de la "demora", condicionado a que todos los participantes estudien en la playa.

En la página web se dispone de una actividad complementaria (capítulo III, apartado 4.2).

Cuando ya tenemos los dos efectos simples representados y estudiados, el siguiente paso es comparar cómo son entre sí los dos efectos simples.

Una manera operativa de hacerse esta pregunta es reformularla como ¿son paralelas las líneas que representan a los efectos simples?

Cuando las líneas que representan los efectos simples no son paralelas, como es el caso de nuestro experimento, se dice que las dos variables independientes interactúan sobre la dependiente.

Como puede ver, otra vez la representación en líneas tiene un papel pedagógico.

Existe *interacción* cuando la manera de actuar de una variable independiente cambia dependiendo del nivel de la otra variable independiente con la que se combina.

En nuestro caso, se olvida más al cabo de una hora, cuando se memoriza en la playa. Éste es el primer efecto simple. En el otro caso pasa algo diferente: cuando se hace bajo el agua, se recuerda igual en el momento que al cabo de una hora. En definitiva, los efectos simples de la variable "demora" son distintos. Por lo tanto, decimos que se ha producido interacción.²⁴

24. Cuando decimos que se ha producido interacción, esto tiene que ser congruente con encontrar en el análisis de datos un efecto de la interacción significativo estadísticamente.

Si las líneas que representan los efectos simples de una variable independiente no son paralelas, entonces nos indican que existe interacción entre las variables independientes.

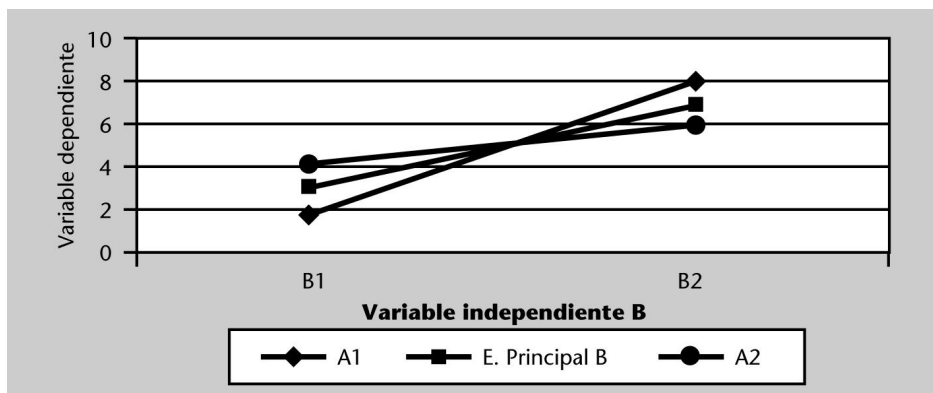
¿Tiene alguna repercusión el hecho de que exista interacción entre las variables? Tiene dos repercusiones:

- Es obligatorio interpretar cada efecto simple por separado, ya que son diferentes.
- Sólo puede interpretarse el efecto promedio –llamado principal– de la variable independiente si se da la circunstancia de que las dos líneas que representan los efectos simples tengan una pendiente del mismo signo. En nuestro caso, ya está suficientemente dicho que no puede interpretarse el efecto principal.

Cuando la línea es horizontal, el signo de la pendiente es cero; cuando sube hacia la derecha, el signo es positivo, y cuando sube hacia la izquierda, el signo es negativo.

¿Puede darse un caso donde haya interacción y el efecto principal pueda interpretarse? Sí, puede darse; veámoslo con un ejemplo gráfico en la Figura 3.7.

Figura 3.7. Caso de interacción con efecto principal interpretable



Para profundizar más sobre los resultados de una investigación factorial, representada gráficamente, puede realizar la actividad 3.4 de la página web.

4.3. La interpretación de los resultados factoriales. Generalización a más variables y a los cuasi experimentos

En el apartado anterior ya hemos expuesto los conceptos de efectos simples, interacción y efecto principal de una variable independiente. Parece intuitivo que al estudiar los resultados de la segunda variable independiente los apliquemos de la misma forma: comenzaremos estudiando cómo son entre sí los efectos simples en la gráfica de la variable "contexto". En nuestro caso, se trata de la Figura 3.6.

Como podemos observar, las líneas no son paralelas y, por lo tanto, deducimos que hay interacción²⁵.

Si se ha detectado interacción en el análisis de la primera variable independiente, necesariamente se detectará interacción en el análisis de la segunda variable independiente.

Puesto que hay interacción, interpretamos cada uno de los efectos simples por separado. En nuestro caso, para el primer efecto simple –línea con rombos–, diremos que memorizar en la playa ayuda al recuerdo si el examen se hace de inmediato; respecto al segundo efecto simple –línea con círculos– diremos que memorizar en la playa hace que se olvide más al cabo de una hora.

En la página web se dispone de una actividad complementaria (capítulo III, apartado 4.3).

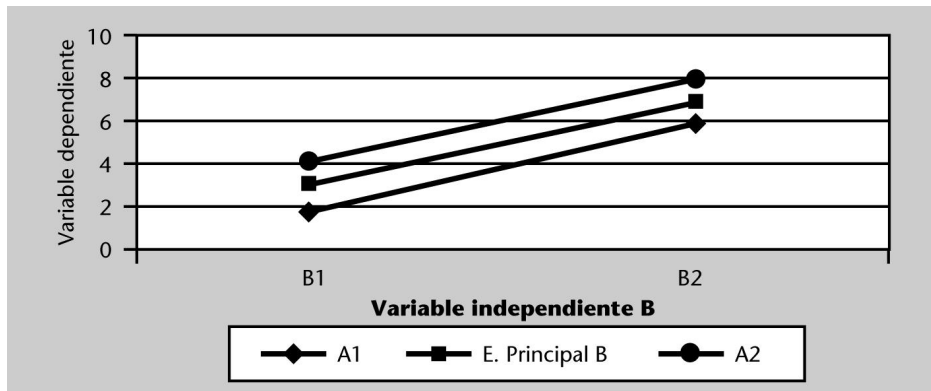
En contra de cómo se presenta en la mayoría de los textos, que empiezan con un caso sin interacción y luego se va complicando hasta que uno deja de leer, nosotros hemos seguido la estrategia opuesta: hemos optado por lanzarle a la piscina primero y, una vez dentro, decir cómo hay que hacer para no hundirse. La razón es que no debe olvidar que siempre se debe empezar por el estudio de la interacción.

Y cuando se encuentre con que hay interacción, porque las líneas de los efectos simples no son paralelas, debe comprobar si la semejanza de las pendientes permite interpretar el efecto principal; porque cuando los efectos simples de una variable independiente no son del mismo signo no se puede interpretar el efecto principal.

25. Si cree que todavía no tiene claro el concepto de interacción y que necesita otra pasada, puede intentarlo con una analogía más sencilla en el caso 3.4 en la página web.

Si ha llegado hasta aquí, ahora le resultará sencillo entender una gráfica donde se presenta una situación sin interacción. Los casos sin interacción son más simples, ya que las variables se comportan de la misma manera, se combinan como se combinan con los niveles de la otra variable.

Figura 3.8. Caso factorial sin interacción



En la página web se dispone de una actividad complementaria (capítulo III, apartado 4.3).

A continuación presentamos una guía resumen para estudiar correctamente todos los resultados obtenidos después de una investigación factorial con dos variables:

1) *Obtener las medias de la variable dependiente en cada una de las (2 x 2) condiciones.*

- Disponer las medias en una tabla de contingencia de 2 x 2.

2) *Realizar una gráfica para cada variable independiente, como variable central (una variable es central cuando se representa en el eje de abscisas).*

3) *Estudiar la primera variable independiente.*

- Comparar las líneas que representan los efectos simples de la variable.

1) Si son paralelas, no hay interacción. Estudiar directamente el efecto principal de la variable.

2) Si no son paralelas, hay interacción. Estudiar cada efecto simple por separado. Se podrá estudiar el efecto principal sólo si son iguales los signos de las pendientes de las líneas que representan los efectos simples.

4) *Estudiar la segunda variable independiente.*

- Comparar las líneas que representan los efectos simples de la variable.
 - 1) Si son paralelas no hay interacción. Debe dar lo mismo que en la primera. Estudiar directamente el efecto principal de la variable.
 - 2) Si no son paralelas, hay interacción. Estudiar cada efecto simple por separado. Podrá estudiarse el efecto principal sólo si son iguales los signos de las pendientes de las líneas que representan los efectos simples.

El caso con que iniciábamos este bloque temático, y sobre el que hemos desarrollado los conceptos, es el más simple posible (2×2).

Si tuviéramos un caso en que una de las variables independientes tuviera 3 niveles y la otra tuviera 4 niveles (3×4), en una gráfica aparecerían 3 líneas con cuatro puntos y en la otra 4 líneas con tres puntos. Estas líneas no tienen por qué ser rectas. Esto significa que el concepto gráfico de pendiente –tan útil para los diseños 2×2 – deberá sustituirse por el de *perfil*.

Además, lo podemos imaginar tan extenso como queramos, por ejemplo, podríamos tener una investigación con el siguiente esquema:

VI_1 (3 niveles); VI_2 (2 niveles); VI_3 (5 niveles);

La manera de indicarlo sería:

se trata de un diseño factorial $3 \times 2 \times 5$

Los diseños factoriales pueden tener variables independientes con un tratamiento intersujetos, intrasujetos o cualquier combinación posible entre ambas. También podemos encontrarnos con factoriales donde todas las variables sean manipuladas, dando lugar a experimentos; o todas no manipuladas, dando lugar a cuasi experimentos, o las posibles mezclas de ambas.

Finalizamos el capítulo con otro experimento factorial, el caso 3.5.

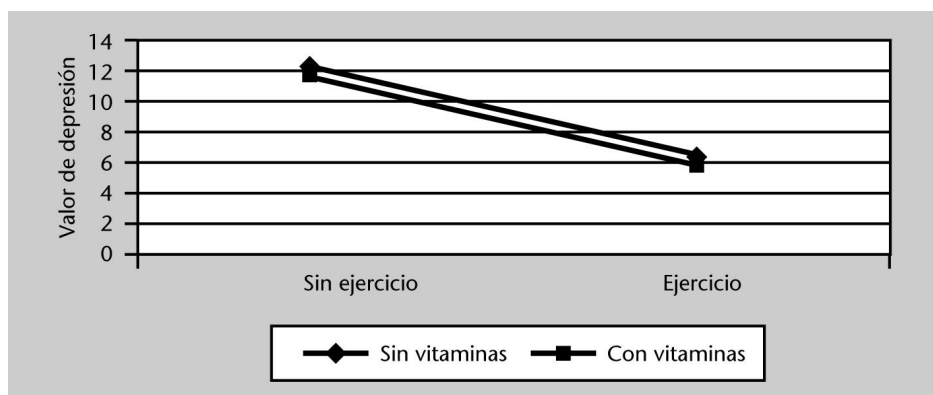
Caso 3.5

Ejercicio físico y depresión

¿La gente más animada es la que hace ejercicio físico? o ¿la gente que hace ejercicio físico es la más animada? Mediante una metodología de investigación con encuestas sólo podemos constatar que se da una relación alta y positiva entre la práctica de ejercicio físico moderado y el estado de ánimo moderado. No podemos saber qué es la causa de qué. Para contestar a esta pregunta necesitamos una metodología experimental. Necesitamos poder manipular alguna de las dos variables. Esto fue lo que hicieron McCann y Holmes (1984). Eligieron manipular el nivel de ejercicio aeróbico. Para que su posible influencia fuera más notable decidieron trabajar con un grupo de chicas universitarias con una leve depresión. Si el ejercicio aeróbico mejora el estado de ánimo, el indicador de depresión mejorará. Adicionalmente,

quisieron probar si un suplemento polivitamínico en la dieta producía elevación del estado de ánimo. Para poner a prueba ambas hipótesis hicieron el diseño experimental siguiente, factorial 2×2 . La variable "grado de ejercicio aeróbico" fue manipulada intersujetos, en grupos aleatorios. De entre el total de las participantes, distribuidas por asignación al azar, un grupo no haría ejercicio (el control) y el otro lo haría (el experimental). La variable "ingesta de vitaminas" fue manipulada intersujetos, en grupos aleatorios. Por lo tanto, la mitad –al azar– del grupo experimental tomaría vitaminas y la otra mitad no. Del mismo modo, –al azar– la mitad del grupo control tomaría vitaminas y la otra mitad no. En la Figura 3.9 presentamos unos datos similares a los encontrados por los investigadores, dos meses después de seguir los distintos planes.

Figura 3.9.

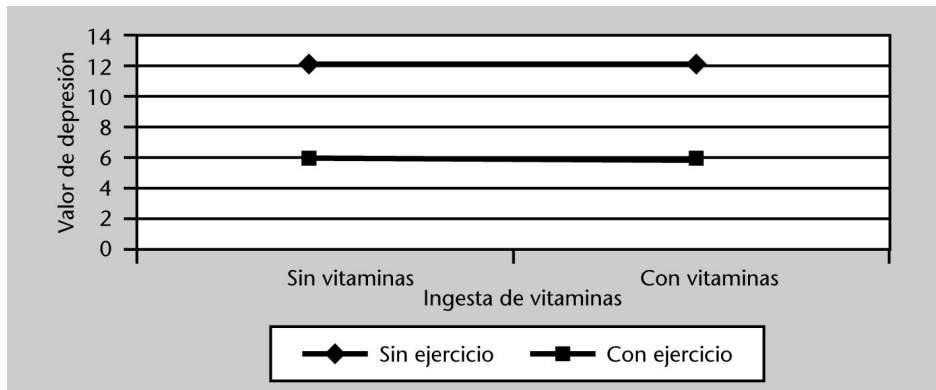


Resultados del ejercicio aeróbico y la ingesta de vitaminas sobre el nivel de depresión de una muestra de chicas, después de sesenta días. Variable central representada: realización de ejercicio aeróbico.

En la Figura 3.9 hemos representado como variable central la realización de ejercicio aeróbico. Como podemos observar, no se produce interacción entre las dos variables: las dos líneas que representan los efectos simples son paralelas. En consecuencia, podemos interpretar el efecto principal de la variable independiente. Podemos concluir que el ejercicio físico aeróbico produce mejora en el estado de ánimo de las participantes, tras sesenta días de realización.

En la Figura 3.10 representamos como variable central la ingesta de vitaminas.

Figura 3.10.



Resultados del ejercicio aeróbico y la ingesta de vitaminas sobre el nivel de depresión de una muestra de chicas, después de sesenta días. Variable central representada: ingesta de vitaminas.

Como podemos observar en la Figura 3.10, ya se trate del grupo que hace ejercicio o del que no hace, las personas que toman vitaminas no se diferencian de las que no las toman, en cuanto a sus valores de depresión.

Capítulo IV

Metodologías cuasi experimentales y selectivas

1. Metodologías cuasi experimentales

1.1. Presentación de un caso: introducción al prefijo *cuasi*

Montse Obach estudió Psicología en la Universidad Autónoma de Barcelona. Siempre tuvo claro que quería dedicarse al ámbito educativo, así que, después de acabar la licenciatura, preparó con ahínco las oposiciones al cuerpo de profesores de la especialidad de Psicología y Pedagogía de la Generalitat. Tuvo suerte y pudo entrar en la lista de interinos el primer año que se presentó. Aunque le costó dos intentos más obtener plaza de funcionaria, ha estado trabajando desde el primer momento. Después de varios años de mucho entusiasmo empieza a sentirse un poco quemada: la lista de tareas que se le asignan parece inabarcable y tiene la sensación de hacer su trabajo muy aislada del resto de los profesores.

En esta situación ha empezado el último curso. En diciembre le llegó un folleto del Colegio de Psicólogos de Cataluña en el que se informaba de la celebración de una jornada con el título de "La salud y el grado de satisfacción del profesorado es clave". Lo que más le atrajo fue una conferencia programada a media mañana con el título de "La motivación del profesorado" que iba a ser impartida por un profesor de la Autónoma de Madrid. Aunque el nombre del conferenciante no le sonaba de nada, el título le gustó y decidió inscribirse. Estaba programada para el 16 de febrero.

Al finalizar ese día, tenía una sensación ambivalente. La conferencia había sido un poco desastre. La actividad programada al principio de la mañana con participación de los partidos políticos se había alargado y había hecho que la conferencia que a ella le interesaba fuera reducida a la mitad y de forma brusca. Así que se quedó frustrada. Sin embargo, como se había presentado por escrito el contenido que estaba previsto impartir, pudo leerlo con tranquilidad en un hueco de la tarde. La lectura resultó fructífera, ya que tuvo la virtud de dejarle una cosa clara: motivar motiva. Y esa perogrullada la puso en acción.

Enseguida pensó en el grupo de diversificación curricular²⁶ al que tenía que impartir Lengua y Matemáticas. Repasó las seis dimensiones de intervención motivacional que se mencionaban en la conferencia (consulte el cuadro 4.1) y empezó a pensar en la manera de ponerlas en práctica en lo que restaba de curso. Se entusiasmó tanto que se imaginaba convenciendo a otros profesores de la bondad de sus esfuerzos. Y esa palabra *convencer* se convirtió en mágica, ya que le abrió la puerta a sus más que enmohecidos conocimientos de metodología de investigación.

Cuadro 4.1

TARGET: dimensiones de intervención motivacional en el aula

La palabra inglesa *target*, que significa 'objetivo' o 'diana', está formada por las iniciales de seis dimensiones de la actividad en el aula con repercusión motivacional: **T** area, **A** utoridad, **R** econocimiento, **G** rupo, **E** valuación y **T** iempo.

T area: se refiere al modo en que se presentan las tareas dentro del aula. Se hace referencia a tres aspectos: la posibilidad de elegir, el reto moderado y la fragmentación de su desarrollo. Frente a la práctica habitual de presentar una única tarea para que todo el grupo trabaje el mismo objetivo, se ha puesto de manifiesto que la posibilidad de elegir entre diferentes tareas genera motivación positiva al implicar un cierto compromiso con la tarea. Además, las tareas que implican un reto moderado son las que generan motivación por el aprendizaje. Las muy fáciles producen aburrimiento, las muy difíciles, frustración. Cuando se plantea fraccionándola en objetivos intermedios se maximiza la probabilidad de que todo el mundo consiga terminar alguna parte con éxito. Y eso es motivador.

A utoridad: los expertos dicen que hay que huir tanto del autoritarismo como de la permisividad. La idea es que la autoridad hay que ganársela y utilizarla de forma flexible. El profesor que lo consigue produce mayor motivación por el aprendizaje que el que no lo hace.

R econocimiento: se refiere al manejo de los elogios por parte del profesor. La cuestión clave para la motivación es el qué y el cómo se elogia. En cuanto al qué elogiar, parece claro que para el desarrollo de una motivación alta, se debe elogiar una manera de funcionar que lleva a desarrollar estrategias para mejorar el aprendizaje. Menos evidente es lo relativo al cómo. La investigación dice que en público debe elogiarse la acción y no a la persona que la realiza. Cuando quiera elogiarse a la persona, es preciso que se haga en privado. Lo mismo hay que decir con respecto a la reprensión.

G rupos: el trabajo cooperativo no sólo genera motivación por el aprendizaje, sino que tiene claros efectos terapéuticos para aquellos estudiantes con baja motivación. Formar parte de un grupo que realiza una tarea con éxito aumenta las probabili-

26. Los grupos de diversificación curricular son una de las medidas que la LOGSE prevé para la atención a la diversidad. Son grupos que se forman con alumnos que tienen dificultades de aprendizaje y a los que se adapta el currículo en su conjunto, adecuándoselo a su nivel. Estos grupos pueden formarse para cursar 3º y 4º de la ESO. Dentro de las funciones asignadas a los especialistas en Psicología y Pedagogía está la de impartir materias instrumentales –como Lengua o Matemáticas– a estos grupos.

dades de aprendizaje de este tipo de alumnos y permite mejorar sus expectativas. Además, en caso de fallar, la responsabilidad es compartida y aumenta la probabilidad de que, entre todos, emerjan soluciones para los fallos.

E *evaluación*: las prácticas clásicas de evaluación –referida a normas, centrada en el producto y de carácter público– fomentan unos tipos de orientación motivacional que pueden tener consecuencias negativas para el aprendizaje. Por el contrario, las prácticas de la evaluación referidas a criterios, centradas en el proceso y de carácter privado fomentan el desarrollo de la motivación por el aprendizaje.

T *tiempo*: se refiere al manejo del tiempo en la actividad escolar. La importancia motivacional de esta dimensión está mediada por su relación con la ansiedad. El tiempo límite produce una tensión que hay que enseñar a manejar de forma positiva –ansiedad facilitadora– y evitar sus potenciales consecuencias negativas –ansiedad inhibidora²⁷.

Convencer la llevó a *demostrar* y esto a *investigar*. ¿Cuál es el mejor plan de investigación para demostrar que puede modificarse la motivación y el rendimiento de los alumnos de diversificación curricular?

Según sus apuntes de *Metodología*, el experimento. ¿Podía hacer ella un experimento con su grupo de alumnos? Sacó papel y lápiz y empezó a poner cosas por escrito. Se dio cuenta de que, en principio, su objetivo era el de un experimento, ya que pretendía contrastar una hipótesis de relación causal: si la estrategia para aumentar la motivación que había leído en la conferencia era eficaz, entonces sus alumnos, al estar más motivados, mejorarían su rendimiento.

Eso quería decir que el rendimiento era la variable dependiente, y la intervención motivacional, la variable independiente. Bueno, debería ser más concreta. Tenía que operativizar la variable independiente. Elegiría sólo algunas de las seis dimensiones de intervención motivacional.

Repasando las dimensiones vio que las que más le atraía aplicar eran el "trabajo en grupos cooperativos" y el "modo de evaluación". Ambas cosas habían pasado por su cabeza en años anteriores, pero la vorágine diaria había hecho que las pospusiera una y otra vez. Anotó que tenía que buscar información para poder preparar las clases de la última evaluación para ser llevadas a cabo mediante el trabajo en grupo. Lo mismo tenía que hacer para adecuar su modo de evaluar el método, cuya eficacia se proponía investigar.

La forma de medir la variable dependiente era fácil: usaría los controles que tenía preparados de otros años. De paso evitaba, al prepararlos a propósito para este curso, hacerlos más fáciles debido a sus expectativas... Caer en la cuenta de que había razonado así de bien le hizo sonreír: ¡con lo que le había costado aprobar los diseños de investigación!

También estaba claro que los participantes iban a ser sus alumnos de diversificación. Pensó en llevar un control exhaustivo de la asistencia para procurar que todos se beneficiaran al máximo de su nuevo modo de trabajo. Si alguno faltaba en ex-

27. Puede consultar Pintrich y Schunk (2002).

ceso, no tendría en cuenta sus resultados. "Otra variable extraña" se oyó decir en alto mientras volvía a sonreír. Como todos ellos iban a recibir el plan de intervención motivacional, el diseño sería intrasujeto. Eso facilitaba también el control de las diferencias individuales.

Tan orgullosa quedó de sus preparativos que, a pesar de ser tarde, llamó a su amiga Ana Soler, compañera de estudios, que había montado con su marido un negocio de turismo rural en un pueblecito del Pirineo leridano. Suelen hablar a menudo para contarse sus respectivos avatares. Ana agradece que le cuente historias del trabajo porque, dice, es su única conexión con la Psicología. Había sido muy buena estudiante y había empezado a colaborar en un proyecto de investigación, pero la vida la llevó por otras latitudes.

En el hostel se acostaban tarde y a Ana no le costó nada contagiarse del entusiasmo de su amiga. Tras escuchar toda la historia y animar a Montse a llevar a cabo su idea, algo apareció en su memoria y quedó callada en mitad de la conversación:

- Ana, ¿sigues ahí? –pregunta Montse, al notar el repentino silencio de su amiga–. ¿Me oyes? –insiste.

- Sí, sí –replica–, es que... –deja en el aire la frase.

- Es que... ¿Qué? –se impacienta la Obach.

- No, nada. Simplemente que me estoy acordando de una cosa... Vamos, que eso que quieres hacer no es un experimento.

- ¿Cómo que no es un experimento?

- Me estoy acordando de que en los casos de aplicación de un tratamiento a un solo grupo, como es el tuyo, no puede hacerse un experimento, porque no pueden establecerse los controles mínimos.

- Pero ¿cómo puedes acordarte de eso? –replica Montse asombrada.

- Porque da la casualidad de que es el mismo caso de la investigación en que participé al año siguiente de terminar; ¿te acuerdas? Iba sobre el tratamiento de la anorexia. Era el mismo. Un grupo de personas con anorexia era tratado con un nuevo sistema que había que demostrar que mejoraba el problema. Dijeron que no era un experimento, sino un cuasi experimento.

- Un cua... ¿qué?

Toda esta historia que hemos inventado trata de ponerle en la situación en que el contraste de una relación causal no puede hacerse en condiciones estrictamente experimentales. Eso suele ocurrir en contextos aplicados como el que se presenta en el ejemplo. Pero el hecho de que no pueda ser un experimento no lo da el contexto, sino la imposibilidad de establecer los controles mínimos.

Cuando la estrategia experimental es de tipo intersujeto, el control mínimo viene determinado por la posibilidad de asignar al azar los participantes a los grupos a los que se aplicarán los diferentes niveles de la variable independiente. Cuando el esquema es intrasujeto, el control mínimo viene determinado por la posibilidad de controlar el orden de presentación de los distintos niveles de la variable independiente.

Cuando quiere contrastarse una relación causal entre variables pero, o bien no pueden asignarse los participantes al azar a los grupos, o bien no puede controlarse el orden de presentación de los niveles de la variable independiente, se utiliza un plan de investigación llamado *cuasi experimental*.

Cuasi es un prefijo que en latín –como puede haber adivinado– significa 'casi'. Es decir, que en el caso mencionado lo que se hace es casi un experimento. Los norteamericanos que acuñaron el término (Campbell y Stanley, 1966; Cook y Campbell, 1979) recurrieron al latín y decidieron llamar *cuasi experimental* lo que en su idioma hubiera sido *almost experimental*. En español, mantenemos el latín para darle un énfasis técnico al término.

1.2. Diseños de medidas antes y después del tratamiento. El grupo de cuasi control

Lo que la protagonista de nuestra historia tenía en la cabeza cuando pensó en demostrar que es posible motivar a sus alumnos del grupo de diversificación curricular era un experimento.

Aunque ya lo hemos anticipado, hay un pequeño detalle que hace que el estudio diseñado por la protagonista de nuestro relato no sea un experimento, sino un cuasi experimento.

En la página web se dispone de una actividad complementaria (capítulo IV, apartado 1.2).

La variable independiente sería la intervención motivacional por medio del trabajo cooperativo y la forma de evaluar. Se aplicaría a base de plantear las actividades en pequeños grupos cooperativos y evaluando el proceso de aprendizaje de cada alumno privadamente.

Compararía dos niveles, sin intervención y con intervención. Ambos niveles se aplicarían al mismo grupo. De hecho, el primer nivel es el que hasta ahora se ha venido aplicando: enseñar matemáticas sin plantear las actividades en grupo y evaluando de modo tradicional.

La variable dependiente es el rendimiento académico en la asignatura en que se interviene: Matemáticas.

En el caso de la aplicación de un tratamiento de mejora –en este ejemplo, de la motivación– no es posible decidir al azar en qué orden se van a aplicar los dos niveles.

Las consecuencias que tiene esta, aparentemente pequeña, pérdida de control experimental son más serias de lo que parece, ya que abren la puerta a la posibilidad de aparición de algunas amenazas para la validez interna.

Tales amenazas no son más que la confusión potencial del efecto del tratamiento con otras variables que, claro está, serían extrañas. Esto hace que, de entrada, el plan de investigación no permita cumplir la tercera de las condiciones para el contraste de una relación causal (capítulo III). Fíjese en que en cualquier experimento puede aparecer una variable extraña. Pero eso se debe a un error en el desarrollo del mismo.

Por muy correctamente que se desarrolle un cuasi experimento, no garantiza la eliminación de las amenazas. Es decir, el problema para la validez interna no está en la aplicación del plan de investigación, sino en el plan de investigación mismo.

Piense en el ejemplo concreto que manejamos. La orientadora/profesora tomará una medida de rendimiento en matemáticas de sus alumnos antes de comenzar el nuevo modo de enseñar. Aplicará su nuevo sistema –más motivador– proponiendo las actividades en pequeños grupos cooperativos y evaluando el proceso de aprendizaje de forma individual y privada. Después volverá a medir el rendimiento en Matemáticas. Si las notas obtenidas después del tratamiento mejoran, pensará que dicho tratamiento ha sido eficaz. Fíjese en que decimos las notas, no el aprendizaje.

Con la enseñanza tradicional, su alumnado obtiene un rendimiento parecido evaluación tras evaluación. Este rendimiento expresa el resultado del proceso de aprendizaje. Lo que espera es que con su nuevo modo de enseñar los estudiantes aprendan más que antes y que ello se refleje en superiores notas.

Este tipo de diseño, que ya sabe que es de naturaleza cuasi experimental, se denomina *diseño pre-post con un solo grupo*.

Su mayor problema consiste en que otras cosas pueden acontecer a la vez que se aplica el tratamiento y que puedan influir sobre el rendimiento. Tales cosas pueden convertirse, por tanto, en variables enmascaradas y constituyen amenazas para la validez interna de la investigación.

Algunas de las amenazas potenciales serían el entusiasmo recuperado de la profesora que se transmite a todo el grupo e influye en el rendimiento independientemente de su nuevo método de enseñanza; se acerca el final de curso y, quien más quien menos, estudia un poco más; se pone de moda una novela para aprender Matemáticas... ¿Hay algo que Montse Obach pueda hacer para descartar estas explicaciones alternativas, para desenmascarar estas variables extrañas?

Una manera muy utilizada de mejorar la validez interna de este diseño es buscar un grupo de comparación. El grupo de comparación del que hablamos sería un grupo al que no aplicamos el tratamiento. En este nuevo esquema de investigación, se mide la variable dependiente en dos grupos, se aplica el tratamiento a uno de ellos y se vuelve a medir la variable dependiente en ambos.

Como ocurría en el caso anterior, esto se parece mucho a un experimento de verdad. ¿Qué hace que no lo sea? Si el investigador puede formar esos dos grupos mediante asignación aleatoria de los participantes de tal modo que estén equilibrados en cuanto a diferencias individuales, la investigación podrá considerarse experimental. Pero, muy a menudo, en los contextos aplicados en que suelen desarrollarse este tipo de investigaciones, los grupos no pueden formarse aleatoriamente. Volvamos a la intervención motivacional para entender por qué.

Montse podría decidir dividir al azar su clase en dos. Pero, además de los inconvenientes logísticos –¿qué hace con una mitad mientras trabaja con la otra si no tiene otra aula ni otro profesor?–, no parece muy ético dejar a la mitad de la clase sin el beneficio potencial del tratamiento pedagógico cuya eficacia se trata de poner a prueba. Parece más lógico buscar un grupo de alumnos lo más parecido posible para que sirva de control, es decir, que estén estudiando lo mismo, pero trabajando con un método tradicional. Para enfatizar que no ha sido formado mediante asignación aleatoria, a este grupo lo llamamos de *cuasi control*²⁸.

El diseño que hemos presentado y que incluye este grupo se denomina *diseño pre-post con grupo de cuasi control*.

Así pues, para mejorar su capacidad de persuasión sobre la eficacia de su intervención motivacional, nuestra protagonista podría pedir colaboración a un compañero profesor de la asignatura de Matemáticas de otro grupo de diversificación con parecidas adaptaciones curriculares. Aplicaría la misma prueba de rendimiento a los dos grupos. Al suyo le aplicaría la intervención motivacional durante la última evaluación, y volverían a hacer un examen idéntico a los dos grupos una vez terminado el periodo de la tercera evaluación. Si la intervención motivacional resultara eficaz, el nivel de rendimiento cambiaría de modo diferente en ambos grupos: la mejora del rendimiento del grupo de Montse, el grupo experimental, sería superior. Ahora bien, si ése fuera el caso, ¿estaríamos seguros de que la diferencia en la evolución del rendimiento de los dos grupos se ha debido al tratamiento?

28. Puede consultar León y Montero (2003).

La sospecha que teníamos con el diseño de un solo grupo de que la mejora podría deberse a que al final del curso siempre se estudia un poco más quedaría eliminada. Esta circunstancia afectaría por igual a los dos grupos y, por tanto, no podría ser responsable de las diferencias. Sin embargo, este diseño no nos permitiría eliminar la posibilidad de que la mejora del grupo con tratamiento se debiera a la motivación producida por el entusiasmo extra de la profesora y su contagio a sus alumnos. Para saber si la variable relativa a la lectura de la novela para aprender Matemáticas podría contaminar la comparación, podría preguntarse explícitamente en los dos grupos por el número de lectores. En la medida en que dicho número fuera parecido, la amenaza desaparecería.

Además del entusiasmo extra, del efecto de la cercanía del final de curso y de la posibilidad de que algún acontecimiento externo incida en la variable dependiente, ahora hay que añadir como amenaza las diferencias individuales que, en principio, no están controladas al no haberse formado los grupos mediante asignación aleatoria. La amenaza más clara es que en el otro grupo haya alumnos muy diferentes de los de Montse. Lo más lógico es que siendo del mismo centro y del mismo tipo, nivel y currículo, ambos grupos se parezcan mucho. Sin embargo, eso hay que comprobarlo. Si se diera el caso, por ejemplo, de que en uno de ellos se hubiera concentrado a los inmigrantes, mayoritariamente magrebíes, debido a que el profesor tutor tiene nociones de árabe, la comparabilidad entre los grupos podría verse mermada.

En cualquier caso, siempre estamos hablando de *amenazas potenciales* a la validez interna; amenazas que, si ocurrieran, darían al traste con el objetivo de la investigación, con la valoración del impacto del tratamiento. El diseño no garantiza la neutralización, pero el investigador puede comprobar si las amenazas han aparecido o no.

En el caso de diferencia grande entre los dos grupos de diversificación curricular, no sería aconsejable utilizar el de los inmigrantes como grupo de cuasi control. Habría que buscar o bien otro grupo o bien otro diseño.

1.3. Otros modos de mejorar el diseño pre-post: otra cohorte, otra variable

El término cohorte hace referencia, fundamentalmente, a un conjunto de personas nacidas en la misma fecha.

En el sistema educativo, una cohorte está compuesta por los nacidos el mismo año natural. La cohorte es utilizada como unidad de gestión del proceso educativo. En las investigaciones de tipo cuasi experimental una cohorte anterior puede ser un buen grupo de cuasi control, ya que, aun teniendo los inconvenientes de no haber sido formada aleatoriamente en relación con el grupo al que se aplica el tratamiento, tiene algunas ventajas frente al grupo de cuasi control sincrónico. Pero vayamos por partes.

En el ejemplo que hemos seguido hasta ahora, el utilizar una cohorte anterior como grupo de cuasi control implicaría que Montse recurriera a sus archivos para tomar el rendimiento en Matemáticas del grupo de diversificación del curso pasado en las dos fechas equivalentes a las que, durante este año, va a evaluar el rendimiento del grupo al que aplica su programa de intervención motivacional. Es decir, si va a tomar la medida "pre" al final de la segunda evaluación y la "post" al final de la tercera, tomará las notas de esas dos mismas evaluaciones del grupo de diversificación al que impartió los mismos contenidos durante el curso pasado. El esquema de investigación es idéntico al anterior, sólo que el grupo de cuasi control es tomado del año anterior. ¿Qué aporta esto a la validez interna?

El efecto de la variable asociada al momento del curso –estudiar un poquito más– sigue controlado. Sin embargo, siguen sin controlarse las variables debidas al posible entusiasmo especial de la "profe" –el año pasado era la misma, pero no le había dado el arrebato investigador– y a la aparición de algún acontecimiento externo a la investigación relacionado con la variable dependiente. ¿Entonces?

La ventaja frente al anterior es que es viable encontrar un grupo de cuasi control equiparado en las variables que permanecen iguales de un año al siguiente.

También sería una buena alternativa en el caso de que sólo hubiera un grupo de diversificación, es decir, cuando no estuviera la posibilidad de encontrar un grupo adecuado de cuasi control para mediar a la vez que el experimental.

Pero la aplicación de este diseño tiene el inconveniente de que no se puede garantizar la igualdad entre las medidas de la variable dependiente.

En nuestro ejemplo, de un año para otro podrían introducirse cambios entre los exámenes que afectarían a esa igualdad. Cuando el cuasi control se efectúa sincrónicamente, el riesgo es más fácilmente controlable.

La otra alternativa al grupo de cuasi control sincrónico es introducir otra variable dependiente en el mismo grupo para que sirva de control intrasujeto. Pero necesitamos una variable que cumpla las condiciones siguientes:

- 1) Poder cambiar de modo no vinculado a los cambios de la variable dependiente del diseño.
- 2) Ser susceptible de ser modificada del mismo modo.
- 3) Tener la misma sensibilidad al tratamiento.

Volvamos al ejemplo. ¿Se le ocurre alguna otra variable dependiente que podamos utilizar y que cumpla esas tres condiciones? ¿La asignatura de Lengua? ¡Pues, claro! El aprendizaje de las Matemáticas no genera aprendizaje en Lengua. Sin embargo, ésta puede enseñarse también utilizando grupos cooperativos y evaluando de la forma mencionada. Además, hipotéticamente, este tipo de tratamiento también podría funcionar con esta otra asignatura. El esquema de investigación será entonces:

- Aplicar pruebas de rendimiento en Matemáticas y Lengua.
- Aplicar la intervención motivacional en las clases de Matemáticas y seguir enseñando igual que antes en las de Lengua.
- Aplicar pruebas de rendimiento en Matemáticas y Lengua.

Este diseño se denomina *pre-post con cuasi control en otra variable*. Si se encuentra la variable que cumpla las condiciones descritas, el diseño tiene algunas ventajas sobre los anteriores.

En la página web se dispone de una actividad complementaria (capítulo IV, apartado 1.3).

La variable relativa a las diferencias individuales desaparece como amenaza, ya que utiliza otro grupo de participantes como grupo de cuasi control. Sin embargo, la variable ligada al entusiasmo extra, en principio, se mantendría. Es imposible hacer que Montse ignore el objetivo de la investigación cuando imparte sus clases en la asignatura en que comienza el tratamiento y cuando lo hace en la asignatura que sirve de cuasi control.

Otra variable que se mantiene como amenaza es la ligada a la igualdad entre exámenes. En este caso es más difícil garantizar que el nivel de dificultad de las pruebas de evaluación sea equivalente. Difícil, pero no imposible. Si nuestra profesora/orientadora ha sido sistemática al analizar las características de las pruebas que haya ido creando puede utilizar criterios de calibración empírica de la dificultad de ambas pruebas que le garanticen su equivalencia.

La dificultad principal para aplicar este diseño está en encontrar otra variable dependiente que pueda medirse de manera similar y que cumpla las tres condiciones que hemos mencionado un poco más arriba.

1.4. Los diseños de serie temporal interrumpida

Vamos a cambiar de problema de investigación para que este nuevo tipo de estrategia quede meridianamente claro. Ya hemos dicho que Montse Obach es a la sazón orientadora de un instituto de enseñanza secundaria. Además de enseñar las materias de Matemáticas y Lengua en el grupo de diversificación curricular, entre sus tareas también está la de coordinar el Plan de Acción Tutorial con los tutores que lo soliciten. Este curso académico, una profesora de Ciencias Sociales le ha sugerido utilizar la hora de tutoría para mejorar los hábitos de estudio de los grupos de primero de ESO con vistas a mejorar el rendimiento en su asignatura. Llevada por su recién descubierto ímpetu investigador, nuestra protagonista decide no sólo aplicar el plan, sino evaluar su eficacia.

Como el curso ya está avanzado, considera que debe empezar con todos los grupos de primero a la vez. Eso hace que, a efectos del diseño, todos formen el grupo experimental, dado que todos reciben el plan de mejora de los hábitos de estudio. Entiende que los hábitos de estudio pueden sufrir muchas variaciones a lo largo del curso, así que se decide por un diseño de serie temporal. ¿Por qué? Siga leyendo y lo verá.

El cambio de estrategia que implica el uso de estos diseños consiste en aumentar el número de medidas que se toman de la variable dependiente antes y después del tratamiento. El diseño más sencillo que usa esta estrategia es el denominado *diseño de serie temporal interrumpida con un solo grupo*.

Su esquema puede resumirse de la forma siguiente:

- 1) Se dispone de un grupo al que se mide repetidamente en una variable dependiente que quiere modificarse.
- 2) La serie de medidas se interrumpe para aplicar el tratamiento.
- 3) Una vez terminado el tratamiento, la serie de mediciones se reanuda.

Veamos cómo quedaría aplicado en nuestro ejemplo.

- Se avisa a los grupos de primero de ESO que hay que enseñar hábitos de estudio.

- Se piden las notas de los controles de la asignatura *Ciencias Sociales* realizados en un periodo de tres meses. (Se han llevado a cabo tres de éstos).
- Se aplica el programa de mejora de los hábitos de estudio durante las horas de tutoría y alguna de la clase de Ciencias Sociales. Esto dura seis sesiones a lo largo de tres semanas.
- Con posterioridad, recoge las notas de los tres controles siguientes.

Llamemos O a cada una de las medidas –por aquello de que constituyen "observaciones"– y utilicemos un subíndice para denotar los diferentes momentos temporales en que se realizan. Si además llamamos X al tratamiento, este diseño se representa así: $O_1 O_2 O_3 X O_4 O_5 O_6$.

¿Qué le parece esta manera de representar los diseños? En el cuadro 4.2 puede ver todos los diseños que le hemos explicado hasta ahora denotados con este sistema.

Cuadro 4.2

Un sistema de notación para los diseños cuasi experimentales

Tiene las indicaciones siguientes:

- 1) Una O para cada observación
- 2) Un subíndice para cada momento temporal
- 3) Una X para indicar la aplicación del tratamiento

Con ellas está perfectamente capacitado para denotar el diseño pre-post con un solo grupo.

¿Le ha resultado difícil? Esperamos a que no. Vamos ahora con el diseño pre-post con grupo de cuasi control. Tenga en cuenta que cada grupo de la investigación debe tener su propia serie de letras y subíndices para dejar constancia de las medidas que en él se han efectuado. Dos grupos, dos series:

Para el grupo con tratamiento: $O_1 X O_2$

Para el grupo de comparación: $O_1 O_2$

Para dejar claro que se trata de dos grupos diferentes, los inventores del sistema de notación (Campbell y Stanley, 1966) decidieron poner una línea entre los dos grupos. Ésta debe ser discontinua cuando los grupos no son formados aleatoriamente. Además, nosotros pensamos que una buena manera de indicar que

el otro grupo aunque no recibe el tratamiento sí participa en la investigación es utilizar una C –de cuasi control– para indicar lo que ocurre en ese grupo mientras el primero recibe el tratamiento. Así que el diseño pre post con grupo de cuasi control se denotaría de la manera siguiente:

$$\mathbf{O}_1 \mathbf{X} \mathbf{O}_2$$

$$\mathbf{O}_1 \mathbf{C} \mathbf{O}_2$$

Para los grupos formados con distintas cohortes se utiliza una línea ondulada. El diseño pre post con control en una cohorte anterior se denota así:

$$\mathbf{O}_1 \mathbf{C} \mathbf{O}_2$$

\/\/\/\/\

$$\mathbf{O}_3 \mathbf{X} \mathbf{O}_4$$

El diseño siguiente es el denominado pre post con cuasi control en otra variable dependiente. Cada serie de medidas –la de variable sobre la que incide el tratamiento (Y) y la de la variable que sirve de cuasi control (Z)– se denota por separado añadiendo otro subíndice para indicar cada una de las variables dependientes. Como sólo es un grupo, no hay línea de separación entre las dos series. El diseño queda así:

$$\mathbf{O}_{1Y} \mathbf{X} \mathbf{O}_{2Y}$$

$$\mathbf{O}_{1Z} \mathbf{C} \mathbf{O}_{2Z}$$

Por último, añadimos la notación del diseño de serie temporal interrumpida para que estén todos los diseños denotados por este sistema dentro de este cuadro:

$$\mathbf{O}_1 \mathbf{O}_2 \mathbf{O}_3 \mathbf{X} \mathbf{O}_4 \mathbf{O}_5 \mathbf{O}_6.$$

Pero nos queda una tarea por hacer antes de acabar el epígrafe. Tenemos que preguntarnos cómo este diseño afronta la aparición de posibles variables extrañas. La inclusión de varias medidas antes y después del tratamiento sirve para desenmascarar amenazas potenciales como son los cambios estacionales o los acontecimientos externos a la investigación.

La serie *pre* reflejaría una posible variabilidad estacional, independiente del tratamiento. La serie *post* reflejaría el posible efecto de un entusiasmo extra de los estudiantes, justo la primera vez que se examinan estudiando con los nuevos métodos (O_4), efecto que quedaría extinguido en las mediciones siguientes (O_5 y O_6). También podemos ver el posible impacto de acontecimientos externos ocurridos en cualquier momento, excepto cuando se aplica el programa.

En el lado negativo para la validez interna sigue estando el hecho de no poder rechazar el posible efecto de las expectativas del investigador.

Como son varios exámenes, hay que vigilar que la dificultad se mantenga constante. Para reposar un poco todas estas reflexiones que hemos ido intercalando sobre posibles problemas con la validez de los diseños que le hemos presentado, hacemos un repaso general que nos permita ir más allá de la casuística de los ejemplos que hemos elegido.

1.5. La validez de los diseños cuasi experimentales

Esperamos que ya le haya quedado claro en capítulos anteriores que cuando quiere concluirse sobre la existencia de una relación de causa-efecto hay que cumplir unas condiciones. Ello se ve facilitado cuando se trabaja en un contexto controlable como es el laboratorio. Sin embargo, en muchas ocasiones la persona que hace la investigación se ve obligada a trabajar en contextos en los que no puede establecer mecanismos de control.

Los diseños cuasi experimentales –y las investigaciones *ex post facto* o selectivas– se desarrollaron para esos casos.

La imposibilidad de establecer los mecanismos de control que mencionábamos al inicio del capítulo –asignación aleatoria, control del orden de presentación– hace que puedan aparecer variables extrañas que amenazan la validez interna –¿los estudiantes mejoran por el plan de intervención motivacional o por el entusiasmo extra que está poniendo la profesora?–. Eso afecta a la seguridad con que establecemos que la VI ha producido cambios en la VD. Es decir, afecta a la validez interna.

Hasta ahora, hemos presentado las amenazas a la validez interna ligadas al ejemplo de la profesora/orientadora y sus planes de investigación. Pero lo que hemos estado haciendo, en realidad, ha sido ponerle ejemplos concretos de amenazas que son generales para todos los casos y, por tanto, tienen nombre genérico, nombre que pasamos a presentar.

En los libros, ya clásicos, de Campbell y Stanley (1966) y de Cook y Campbell (1979), se mencionan siete amenazas potenciales en los diseños cuasi experimentales –que, en algunos casos, pueden serlo también en los experimentales:

- 1) La historia
- 2) La maduración
- 3) La adaptación a las pruebas
- 4) La instrumentación
- 5) La regresión a la media
- 6) La pérdida no aleatoria de sujetos
- 7) La selección de muestras

Además, existe la posibilidad de que las siete actúen mediante efectos conjuntos –en interacción– unas con otras. Las interacciones más probables pueden darse entre la selección y alguna de las demás.

La *amenaza por efecto de la historia* se produce cuando un acontecimiento, fuera del control de la investigación, y compite con el tratamiento como explicación de los cambios observados en la muestra.

En el caso de Montse Obach, el impacto de la lectura de una novela para aprender Matemáticas sería un caso de amenaza por efecto de la historia.

El hecho de estudiar algo más porque se acerca el final del curso puede entenderse como un caso particular de amenaza por el efecto de maduración. En realidad es una variable ligada al funcionamiento del organismo bajo observación que puede competir con la variable independiente. El prototipo de efecto de la maduración estaría implícito en la expresión de "el tiempo lo cura todo".

La *amenaza de la maduración* consiste en que, cuando el tratamiento se alarga, no sabemos si lo que produce los cambios es el tratamiento o éstos son fruto de cambios internos, propios del organismo.

De las *amenazas de adaptación a las pruebas y de instrumentación* ya hemos hablado explícitamente.

La primera es otro modo de llamar al efecto de la práctica que vimos en los diseños intrasujeto, mientras que la instrumentación aparece ligada a que las técnicas de medida pueden sufrir cambios que se confunden con el efecto del tratamiento.

La regresión a la media es una amenaza de la que no hemos hablado en los ejemplos.

La *regresión a la media* es una tendencia natural de puntuaciones extremas a regresar a sus valores medios, cuando se repite la medición.

Un ejemplo claro sería suspender un examen por haber tenido un mal día. Cuando eso ocurre, uno espera que de repetirlo, sin estudiar más, de forma natural, sacaría una nota más cercana a su media de rendimiento.

Este fenómeno de la regresión a la media se convierte en amenaza para aquellas investigaciones que utilizan el rendimiento en una prueba como criterio para formar los grupos.

Por ejemplo, si utilizáramos un examen para decidir quién hace el plan de mejora de los hábitos de estudio, algunos de los seleccionados lo serían por haber tenido un mal día, no por tener un bajo conocimiento de la materia. Al repetir el examen después del tratamiento mejorarán su rendimiento independientemente de la eficacia del mismo. Mejorarían, simplemente, porque están regresando a su media.

Cuando hay que seleccionar casos extremos para hacer una investigación, la manera de conjurar la amenaza es haciendo varias pruebas de selección y utilizar el conjunto como criterio clasificatorio. En nuestro ejemplo, no hemos mencionado la regresión a la media porque es poco verosímil que se tome la decisión de incluir a un estudiante en el grupo de diversificación por su rendimiento en una única prueba. Fíjese en que los miembros del grupo de diversificación son casos extremos de rendimiento académico –por lo que podría aparecer el efecto de la regresión a la media–, pero su rendimiento es reiteradamente medido.

Cuando se toma la decisión de incluir a alguien en un grupo de estas características no se hace con la información obtenida en un mal día.

La *pérdida no aleatoria de sujetos* podría haberse producido en nuestro ejemplo. ¿Recuerda el relato inicial? En él se describía la reflexión de Montse en el proceso de elaboración de su estudio y veíamos cómo tomaba la decisión de no tener en cuenta la evolución de aquellos estudiantes que faltaran muy reiteradamente. Si fuera el caso, si varios de los estudiantes dejaran de acudir a clase, estaríamos ante una situación típica de pérdida de sujetos que podría no ser aleatoria y, por tanto, contaminar la investigación. Para ello debería darse la circunstancia de que las razones del abandono estuvieran relacionadas con la aplicación de la intervención motivacional. Si los menos motivados rechazan el plan y dejan de ir a clase y si el resto mejora, la profesora podría concluir erróneamente que su intervención motivacional ha sido eficaz.

En la página web se dispone de una actividad complementaria (capítulo IV, apartado 1.5).

Por último, cuando se selecciona a los participantes en la investigación por poseer alguna característica, pueden aparecer otras ligadas a ésta que compitan con la variable independiente a la hora de explicar los cambios en la dependiente. Eso se conoce como *sesgo de selección*.

No parece que haya sido el caso en el ejemplo de la intervención motivacional. Pero sí puede aparecer una interacción entre selección e historia. ¿Cómo? Cuando decimos que es necesario comprobar cuántos alumnos de cada grupo han leído la novela que sirve para aprender Matemáticas, estamos queriendo cerciorarnos de que no puede darse tal interacción. Si el número de lectores fuera distinto, el efecto de la historia aparecería de forma diferencial para los dos grupos, para las dos muestras seleccionadas.

Del mismo modo que nos hemos inventado una historia para que entienda la lógica de un tipo determinado de diseños de investigación, a continuación nos servimos de una historia real para introducir la lógica del grupo de diseños siguiente.

2. Diseños ex post facto

2.1. Una historia real. Un caso de investigación *ex post facto*

"A pesar de haber transcurrido más de veinte años, Begoña no ha olvidado lo que ocurrió aquella tarde de mediados del mes de mayo. Por aquel tiempo, estaba en el tercer curso de sus estudios universitarios. Tras pasar la mañana en la facultad, solía comer en casa con sus padres y sus dos hermanos varones. Gonzalo, el mayor, tenía entonces veintidós años, estaba haciendo la mili y disfrutaba de un pase *per-nocta* que le permitía llegar a casa al mediodía. El otro hermano, año y pico más joven que ella, también era estudiante universitario y tenía un horario parecido. Sus tres hermanas estaban todavía estudiando en el colegio y volvían a casa media tarde.

La vida de Begoña discurría normalmente entre sus estudios, sus amistades y su pasión por los caballos, de los que disfrutaba en sus escapadas por tierras extremeñas. De las cosas que guarda recuerdo de aquella época, le gusta recrear las conversaciones con su hermano mayor mientras paseaban a la caída de la tarde. Gonzalo era un chico más bien introvertido y existencialista. El cumplimiento del servicio militar se le hacía cuesta arriba. La actividad y valores militares no eran santo de su de-

voción. Además, su vida sentimental pasaba por sus peores momentos. Su amiga más íntima, de la que siempre había estado enamorado, se casaba con su novio de toda la vida y le había pedido que dejaran de verse.

A Begoña le gustaba oír las confidencias de su hermano, aunque no siempre pudiera animarlo. De las que le hizo aquella primavera, hubo una que recuerda con nitidez:

-Begoña –preguntó Gonzalo, mientras paseaban–, si me pasara algo, ¿tú qué pensarías?, ¿cómo lo vivirías?

-¡Haz el favor de no decir gilipollices!

Y continuaron andando en silencio.

Fue pocos días después de esta breve conversación cuando, estando en casa por la tarde, a la hora en que su padre dormía la siesta en el salón, se oyó un estampido. Begoña salió de su habitación y se encontró con la mirada de susto e interrogación de su hermano pequeño. La casa de sus padres era de esas antiguas, con techos altos y pasillos largos. Cuando los dos se miraron, nada se oía a ninguno de los lados del pasillo. El hermano reaccionó primero y se dirigió corriendo al cuarto de Gonzalo, llamándolo.

La puerta está cerrada con pestillo. Al ver que no puede abrir, toma impulso y la desencana de una patada. Entran los dos, ella detrás. Lo que ve le hace gritar. Gonzalo está en el suelo, boca arriba, sobre un colchón que él mismo ha preparado. Tiene el pecho destrozado por el impacto del disparo de su escopeta de caza que ha atado a la cama. Su hermano trata de reanimarlo pero está prácticamente muerto. Mientras tanto, la madre, que descansa en su cuarto, ha oído los gritos y ha salido al pasillo. Cuando se da cuenta de que algo grave ha pasado, se acerca y entra también en la habitación de Gonzalo. Conmocionada, se derrumba y abraza el cuerpo de su hijo, llorando. No se separa de él en un largo rato.

El hermano menor de Begoña va hasta el salón donde el padre sigue durmiendo la siesta. Lo despierta y le da la noticia a bocajarro: "Gonzalo está muerto, se ha pegado un tiro". El padre es incapaz de reaccionar. Padece del corazón desde hace tiempo. Pide a su hijo que llame a su médico. Se queda postrado el resto de la tarde. Entretanto, Begoña llama por teléfono a algunos familiares. Su tío, el militar, le da instrucciones estrictas: "No lo mováis, no toquéis nada hasta que lleguen el juez y la policía militar". Intenta también localizar a un sacerdote. El de su parroquia no está. Después llama a sus tías, hermanas solteras de su padre. Tratan de tranquilizarla. Saldrán para su casa y llevarán a un sacerdote amigo.

Transcurre el tiempo y el susto da paso a la desolación. La madre sigue llorando abrazada al cuerpo muerto de su hijo. Ni Begoña ni su hermano saben qué hacer para consolarla. Por fin, alrededor de las cinco, llegan las tías junto con el sacerdote. La situación es dramática. Todos, excepto la madre, están de pie en la habitación, en corro en torno a Gonzalo. El sacerdote abre su Biblia y encuentra un pasaje del Evangelio de Juan. Su voz lenta y sosegada se va imponiendo al rumor de los llantos:

-No se turbe vuestro corazón. En la casa de mi Padre hay muchas moradas; si no, os lo habría dicho. Voy a prepararos un lugar...

Se hace el silencio. La madre de Gonzalo, por fin, se levanta y se sienta en la única silla del cuarto. A Begoña le viene a la cabeza aquella pregunta: "si me pasara algo, ¿tú qué pensarías?".

Este texto lo hemos tomado de León y Montero (2003; pp. 300-301).

Como ya le habíamos señalado, esta historia aconteció realmente –en Madrid, a principios de los ochenta– y no es la primera vez que la contamos. La impotencia que generan determinadas experiencias puede derivar, fácilmente, en culpabilidad, tal y como trasluce la pregunta que resuena en la cabeza de la testigo de nuestro relato. Pero el estado actual de conocimientos sobre las causas de la conducta suicida muestra que es muy difícil predecir un acontecimiento como la muerte de Gonzalo.

¿Es que los investigadores que se han dedicado a ese tema han sido especialmente ineptos? Teniendo en cuenta que alguien de la talla de Émile Durkheim –que escribió el famoso tratado sobre el suicidio hace más de un siglo (Durkheim, 1897)– ha sido uno de los investigadores pioneros, no parece que por esa línea argumental vayamos a encontrar la respuesta sobre el relativo desconocimiento de las causas del suicidio.

La dificultad no está en las habilidades de los investigadores, sino en las condiciones en que trabajan, condiciones que, en este caso, vienen determinadas por la naturaleza del problema de investigación. Después de haber estudiado el capítulo III y la primera parte de éste, tendrá claro que la mejor manera de poner a prueba una hipótesis de relación causal es llevando a cabo un experimento. Pero si sospechamos que algo puede causar el suicidio de los seres humanos, no lo manipulamos en el laboratorio con la intención de que en el grupo experimental haya más suicidas que en el grupo de control.

En el caso que nos ocupa, el investigador entra en escena cuando el suicidio ya ha acontecido. Eso en latín se decía *ex post facto* –después del hecho– y esa expresión da nombre a un conjunto de diseños cuya lógica vamos a presentar a continuación.

2.2. Estrategias retrospectivas

Cuando tenemos que investigar sobre algo que ya ha acontecido –si tenemos que investigar *ex post facto*– pueden darse dos situaciones.

a) Lo que ya ha pasado ha sido toda la cadena: han aparecido las causas y han producido el efecto. Además, el investigador se ve obligado a empezar estudiando la variable dependiente. Por ejemplo: Variable dependiente: abandono de la ESO; variable independiente: desconocida.

b) Las causas ya se han dado, pero el efecto está por llegar. Por ejemplo: Variable independiente: valores en CI; variable dependiente: rendimiento académico tres meses después.

En el caso a), el investigador sólo puede tratar de reconstruir los acontecimientos, necesita empezar con la dependiente, ya que no tiene conocimiento de cuál puede ser la independiente.

En el caso b), tiene una independiente candidata que ya se ha dado y una dependiente que se dará y podrá medirla.

Del primer caso nos ocupamos en este epígrafe de estrategias retrospectivas.

El segundo caso da lugar a las estrategias prospectivas que están recogidas en el siguiente epígrafe.

La característica fundamental de la estrategia retrospectiva, desde el punto de vista de cómo se procede en la investigación, es que se empieza seleccionando a los participantes por sus puntuaciones en la variable dependiente y, hacia atrás, se buscan potenciales variables independientes.

En estas condiciones trabajan todos los investigadores que se preguntan sobre la causa de cualquier problema o trastorno en Psicología. Aunque el suicidio puede ser el caso más dramático, piense en la investigación sobre las causas de determinados trastornos psicopatológicos como la depresión o la esquizofrenia, las causas del fracaso escolar o las del absentismo laboral.

Vamos a centrarnos en la conducta suicida. Cualquiera que quiera investigar sobre sus causas tendrá que examinar con detenimiento aquellos casos en los que se hayan dado suicidios o, al menos, intentos de suicidio. Imaginemos que nos decidimos por los suicidios consumados. ¿Qué podemos hacer como investigadores una vez que ya hemos localizado los casos? Como psicólogos que somos, vamos a preguntarnos, y preguntar a las personas que hayan vivido cerca de las personas suicidas, por las circunstancias que han rodeado el hecho y por las características de la víctima. ¿Ha habido algún acontecimiento estresante reciente? ¿Tomaba algún medicamento o droga? ¿Cómo eran sus relaciones personales y familiares? ¿Cómo le iba últimamente en el trabajo o en los estudios? ¿Padecía algún tipo de trastorno psicológico? ¿Y alguien de su familia? ¿Había antecedentes de intentos autolíticos? ¿Algún antecedente familiar de suicidio, consumado o intentado? ¿Ha dejado algún tipo de mensaje explícito o implícito?

Podríamos seguir formulando preguntas, pero estas pocas que se nos ocurren a bote pronto sirven para ilustrar la cantidad de variables independientes potenciales que podemos postular como candidatas para tratar de explicar lo acontecido. Dese cuenta de que la decisión de hacer una pregunta –incluir una variable en la búsqueda de causas potenciales– tiene algo de arbitraria.

El investigador, guiado por su intuición o por los conocimientos previos en su campo, decide buscar en una dirección determinada, pero al hacerlo asume el riesgo de que otra posible variable independiente quede fuera de la investigación.

A ese peligro, que afecta a la validez de la investigación, se le llama *sesgo retrospectivo*.

Su mayor inconveniente es que no hay modo, *a priori*, de establecer su ocurrencia.

Asumido ese riesgo y medidas las variables elegidas, ¿qué hacer para poder tomar decisiones acerca de la hipótesis de que, efectivamente, tales variables son las que han estado implicadas como causas del fenómeno que estamos analizando?

Nuestra única pista es el análisis de la coincidencia o la covariación entre los valores de la variable dependiente y los de las potenciales variables independientes. Será mejor candidata a variable independiente aquella que más coincida o correlacione con la variable dependiente. También es posible que las correlaciones pongan de manifiesto que hay una combinación de causas potenciales, que es la que mejor clasifica los casos observados. Las variables con altas correlaciones serán mantenidas como potenciales independientes, y las que tengan baja o nula correlación serán desechadas como tales.

¿Podemos estar seguros de que las variables con una alta correlación son la causa que estábamos buscando? Recordará que la covariación no implica causalación. Eso es, entre otras razones, debido a que la relación observada puede ser espuria. Es decir, dos variables que en realidad no tienen relación aparecen covariando al estar ambas relacionadas con una tercera.

Supongamos que hemos llevado a cabo una investigación sobre suicidio consumado y hemos evaluado en un grupo de suicidas un conjunto de variables que consideramos sus causas potenciales. Medimos las mismas variables en otro grupo de personas similares, no suicidas, en cuanto a características sociodemográficas pero sin ningún tipo de antecedente suicida. A continuación unimos los dos grupos de personas. En la variable "Suicidio", los primeros tendrán un 1 y los segundos un 0. Imaginemos que hemos obtenido las correlaciones que aparecen en la tabla 4.1.

Tabla 4.1. Correlaciones entre variables independientes y suicidio

	Acontecer estresante	Depresión	Antecedentes familiares	Conducta antisocial	Malos tratos	Tratamiento psiquiátrico	Consumo de drogas	Éxito laboral
Suicidio	.12	.15	.57	.64	.13	.41	.34	-.32

Aparecen, por un lado, variables con baja correlación y, por otro, variables con una correlación moderada.

Suponiendo que el valor crítico para alcanzar la significación estadística en el conjunto de los casos analizado estuviera en .22, concluiríamos que ni los acontecimientos estresantes ni la depresión ni los malos tratos sufridos en la infancia pueden ser consideradas como causantes de los suicidios observados en la muestra. Eso no quiere decir que ningún suicida de los estudiados padeciese depresión, sino que de los que la padecían los había tanto entre los sujetos suicidas como entre los no suicidas.

Ahora bien, ¿podemos afirmar con seguridad que, por ejemplo, los antecedentes familiares y la conducta antisocial son causas de suicidio? Desgraciadamente, no. El hecho de que dos personas pertenecientes a la misma familia tomen la decisión de suicidarse puede ser explicado por el hecho de que compartan algún tipo de característica disposicional o circunstancial. Tales características, por el momento desconocidas, serían la causa de los dos suicidios y por eso éstos aparecen relacionados. Algo parecido podría postularse de la conducta antisocial. Si se considera el suicidio como un tipo de agresividad autorreferida, desde una perspectiva bioquímica podría postularse que determinados trastornos en la producción de los neurotransmisores que regulan la agresividad producen tanto la conducta antisocial –agresividad hacia otros– como la conducta suicida –la agresividad hacia sí mismo.

Fíjese, además, en que no hemos incluido en nuestra investigación ninguna medida de los aspectos relativos a la producción de tales neurotransmisores.

Quizás le parezca que en estas condiciones no merece la pena investigar. Tenga en cuenta que en muchos casos es el único modo que tenemos para avanzar algo en nuestro conocimiento de problemas de alta relevancia social.

¿Es posible mejorar las condiciones para el contraste de las hipótesis causales cuando investigamos este tipo de problemas? Hay dos vías potenciales.

La primera ya la conoce. Una vez detectadas las posibles causas del problema investigado mediante una estrategia de tipo retrospectivo, éstas pasan a ser postuladas como factores de riesgo. Si se conoce algún procedimiento para

paliar tales riesgos, se prepararán programas sistemáticos de prevención. La valoración de la eficacia de su aplicación, mediante diseños cuasi experimentales, será una vía indirecta de contraste de la hipótesis causal.

La otra vía consiste en tratar de pasar de un esquema retrospectivo a otro prospectivo.

Pero eso no siempre es posible.

En el cuadro 4.3 se presenta una investigación llevada a cabo con pacientes atendidos en urgencias psiquiátricas por intentos autolíticos, en la que se sigue una estrategia retrospectiva para poner a prueba la importancia de factores disposicionales como la agresividad y la impulsividad en la conducta suicida. Al final se considera la posibilidad de seguir la investigación con una estrategia prospectiva.

Cuadro 4.3

Agresividad, impulsividad e intentos de suicidio

Un grupo de investigadores madrileños (Rubio, Montero, Jáuregui, Marín y Santo-Domingo, 1998) trabajaron en un servicio de Psiquiatría recogiendo 232 casos de intentos de suicidio. Para aumentar la calidad de su investigación mediante el aumento de la heterogeneidad y tamaño de la muestra, recogieron también información relativa a 116 personas con diagnósticos que implican alto nivel de agresividad. Por último, incluyeron a 324 personas sin ningún tipo de problema psiquiátrico.

A todos les fueron aplicadas tres escalas: escala de violencia (EV), escala de impulsividad (EI) y escala de riesgo suicida (RS), desarrolladas en Estados Unidos por Plutchik y Van Praag (1989) y traducidas y adaptadas a la población española por el mismo equipo de investigación. Además, se recogió información sobre datos sociodemográficos, acontecimientos estresantes, antecedentes psiquiátricos personales y familiares.

A partir de las correlaciones de todas estas medidas con la respuesta al último ítem de la escala RS, "haber realizado alguna tentativa de suicidio a lo largo de su vida", se identificó la mejor combinación de variables a la hora de clasificar la muestra entre los que contestaban afirmativamente al mencionado ítem y los que lo hacían negativamente. Se consiguió un porcentaje del 90% de clasificaciones correctas a partir de cinco variables: riesgo suicida (.91), antecedente personal de diagnóstico psiquiátrico grave (.53), impulsividad (.46), violencia (.43) y edad (-.18). Los números entre paréntesis expresan en términos de correlación la importancia de cada variable en la combinación final.

¿Podrían predecirse nuevos intentos de estos datos? Los autores señalan que "algunas variables que están presentes en el momento de la tentativa pueden modificarse durante el seguimiento como consecuencia de iniciar un tratamiento... o de los cambios relacionales posteriores que pueden disminuir la probabilidad de que la tentativa se repita" (Rubio *et al*, 1998, p. 325).

2.3. Estrategias prospectivas

- 1) ¿Le gusta salir a menudo?
- 2) ¿Daría dinero para fines caritativos?
- 3) ¿Prefiere leer a conocer gente?
- 4) ¿Frecuentemente improvisa decisiones en función de la situación?

Si su respuesta es afirmativa a las preguntas primera, segunda y cuarta, y negativa a la tercera, tiene algunas características personales que los psicólogos denominan *extraversión*. Si sus respuestas van justo en el sentido contrario, esas características responden a la personalidad *introvertida*.

La dimensión bipolar de la personalidad²⁹ denominada *introversión-extraversión* es una de las que más tradición tiene dentro del campo de la Psicología que se ha preocupado por estudiar dimensiones estables en el comportamiento de la gente. El modo de medir tales dimensiones es el uso de cuestionarios estandarizados. Las preguntas que le acabamos de formular no son más que algunos de los ítems que componen la escala E dentro del cuestionario EPQ-R, cuyas siglas responden en inglés a 'cuestionario de personalidad de Eysenk, revisado' (Eysenk y Eysenk, 1985).

Aunque bien es cierto que en la Psicología se ha discutido mucho sobre la pertinencia de postular dimensiones de personalidad estables comunes a todos los individuos, nuestro interés al traerlas a colación tiene que ver con las peculiaridades de los diseños de investigación que trabajan con ellas como variables independientes. Es decir, nos interesa ilustrar las peculiaridades de investigaciones que quieran poner de manifiesto que la personalidad influye sobre otra variable³⁰.

29. Esta dimensión bipolar de la personalidad ha sido postulada por los psicólogos que tratan de explicar la conducta humana como fruto de predisposiciones estables de los individuos. Las personas extravertidas son sociables, vivaces, activas, asertivas, buscadoras de sensaciones, despreocupadas, dominantes y aventureras. En el polo opuesto se encuentran las personas introvertidas.

30. Puede consultar Hernández (2000), para una revisión reciente.

Veamos un ejemplo. Los historiadores de la educación sostienen que el origen de la palabra *universidad* –del latín *universitas*–, hace referencia al hecho de que en las primeras escuelas catedráticas que cobraron fama al final de la Edad Media, como la de París o la de Bolonia, se organizaban los estudiantes por sus lugares de procedencia. Al extenderse la importancia de tales escuelas, el lugar que acogía a los estudiantes era una suerte de muestra universal, al ser cada vez mayor la variedad de sus orígenes. De algún modo, acudir a la Universidad suponía entrar en contacto con todo el mundo conocido (Bowen, 1990). Parece que unos siglos más tarde es posible acceder a todos los rincones de nuestro planeta sin movernos de nuestra habitación. También es posible cursar estudios universitarios de ese modo. ¿No le parecen experiencias tremendamente parecidas y, a la vez, tremendamente dispares?

Pero volvamos a lo nuestro. Todo este preámbulo viene al caso para que tome conciencia de que las características personales de los estudiantes medievales y los que actualmente estudian a distancia a través de redes telemáticas podrían ser muy diferentes. Imagine que es el responsable de la política de promoción de una universidad a distancia que trata de ser pionera en el uso de redes para desarrollar su actividad y se empieza a correr la voz de que este sistema no funciona para personas con un alto nivel de extraversión. O dicho de otro modo, que el tipo de personalidad del estudiante es un factor que influye de forma muy importante en el rendimiento a distancia.

Como tiene una buena formación metodológica, la primera medida que toma es la de contrastar la veracidad del rumor. Se le ocurre lo siguiente:

- 1) Incluir una escala de personalidad que mida "extraversión" dentro de los datos que se requieren en el momento de la matrícula.
- 2) Elegir dos grupos, de igual tamaño, de estudiantes extravertidos e introvertidos.
- 3) Comparar el rendimiento al final de curso.

Como medidas de control, decide incluir un número idéntico de estudiantes de cada titulación, de cada curso y de cada sexo. Es decir, decide controlar estas tres variables extrañas.

En la página web se dispone de una actividad complementaria (capítulo IV, apartado 2.3).

Fíjese en que, con respecto a la investigación retrospectiva, hemos cambiado el procedimiento: primero se seleccionan personas con determinados valores en la variable independiente –la personalidad– y después se espera –en otros casos, se provoca– la aparición de los valores en la variable dependiente –el rendimiento.

El investigador mira hacia adelante, trabaja de forma prospectiva. Este diseño es *ex post facto prospectivo*.

¿Por qué no se considera un experimento este estudio? Porque el investigador no puede manipular la variable independiente, sólo puede seleccionar personas que ya poseen un determinado valor en la misma. ¿Y qué importancia tiene? Ese pequeño detalle, esa diferencia entre manipular y seleccionar, tiene importantes consecuencias para el control de variables extrañas.

El riesgo fundamental en el diseño prospectivo es el *sesgo de selección*.

Éste consiste en que el investigador no puede separar la característica seleccionada –el valor de la variable independiente– de otras relacionadas con ellas y que podrían a su vez influir en la variable dependiente.

Si no puede separarla, no puede controlar las otras características que pasan a ser variables enmascaradas potenciales.

En nuestro ejemplo podría darse la circunstancia de que, efectivamente, los datos pusieran de manifiesto que los introvertidos rendían mejor que los extrvertidos. Si su Universidad es de carácter público, se encuentra en un atolladero, porque debe procurar la igualdad de oportunidades de los ciudadanos, más allá de sus características personales.

¿Cree que el sesgo de selección puede haber contaminado los resultados? Es decir, además de la posibilidad lógica de que hubiera alguna variable enmascarada detrás de la introversión, ¿se le ocurre alguna en concreto que tenga visos de verosimilitud? Está claro que ni la titulación ni el curso ni el sexo han podido enmascararse, ya que se tomaron medidas al respecto. ¿Podría ser que los más capacitados para el rendimiento en tareas académicas fueran los más introvertidos? Que sepamos, no hay relaciones claras entre personalidad y aptitudes, consideradas éstas de un modo general.

Reunido con su consejo de sabios y después de dar muchas vueltas al asunto, empieza a aparecer una posibilidad que le permitiría salvar la situación. Comentando los datos con algunos profesores, alguien señala que hay estudiantes que sistemáticamente consultan mucho, y otros, nada. ¿Podría eso tener que ver con la personalidad? Parece que sí, pero ¿cómo puede haber influido en el rendimiento...? ¡Hey! ¡Una luz aparece en su cabeza. Descuelga el teléfono y llama a su centro de proceso de datos. Pregunta si es posible dividir a los estudiantes, además de por su personalidad, por el hecho de que las asignaturas sean evaluadas, o no, de forma anónima. Le responden que aunque

no tienen la información de todos los casos, pueden conseguirla en unos días. La espera resulta productiva y le entregan los datos que aparecen en la Tabla 4.2.

Tabla 4.2. Rendimiento medio según personalidad y tipo de evaluación

	Extravertidos	Introvertidos	Media
Evaluación anónima	8	8	8
Evaluación no anónima	10	8	9
Media	9	8	

El análisis de los datos revela que entre los estudiantes evaluados de forma anónima no hay diferencias de rendimiento en función de la personalidad: introvertidos y extravertidos obtienen la misma media (8). Sin embargo, cuando la evaluación no es anónima, es decir, que el profesor conoce a los alumnos, los extravertidos obtienen mejores notas (10 frente a 8). Queda por evaluar si, como sospechamos, los extravertidos son los que hacen las consultas (y a los que conoce el profesor). Suponiendo que esto sea así, significaría que los extravertidos han sido más generosamente evaluados cuando el procedimiento no era anónimo y, una razón verosímil, podría ser que los profesores generaran una imagen positiva que se reflejó en las notas. Conclusión: es mejor hacer evaluaciones anónimas.

Fíjese otra vez en la Tabla 4.2. ¿No le resulta familiar? ¡Exacto! Tiene el mismo aspecto que la de los resultados de un diseño experimental factorial 2 x 2. En este caso no lo es, porque ninguna de las dos variables ha sido manipulada.

¿Podemos calcular la interacción entre dos variables no manipuladas? Sin ningún problema.

Nuestro problema como investigadores cuando estamos trabajando en un contexto *ex post facto* es que estamos menos seguros de que esa interacción que observemos sea entre las variables estudiadas y no entre otras, posiblemente enmascaradas.

Una última apreciación. ¿Podríamos haber manipulado la variable "tipo de evaluación"? Sin ningún problema. Podríamos diseñar un estudio en el que los participantes seleccionados por su personalidad fueran asignados aleatoriamente a dos modos de evaluación: anónima y no anónima. Si sacamos a colación este asunto es para que caiga en la cuenta de que pueden diseñarse

estudios en los que se trabaje con variables manipuladas –lo cual les da rango de verdadero experimento– y con variables no manipuladas –lo cual tiene rango de estudio *ex post facto*.

Para no complicar la terminología con que nos referimos a cada tipo de estudio, nuestra propuesta es llamar a este tipo de estudios diseños factoriales –por tener más de una variable independiente– y después denominar a cada una de las variables según cómo se haya llevado a efecto su inclusión.

Así, podríamos tener un diseño factorial con dos variables en el que ambas fueran experimentales, ambas fueran *ex post facto* o fuera una de cada tipo. Saber eso aclara las condiciones de validez interna de cada una de las hipótesis que pueden contrastarse en este tipo de estudios.

3. Diseños evolutivos

3.1. La edad como variable independiente

El término *diseños evolutivos* hace referencia a aquellas investigaciones en las que la edad –o el paso del tiempo– se estudia como variable independiente³¹.

La disciplina que más los ha utilizado ha sido, lógicamente, la Psicología del Desarrollo. Eso no quiere decir que los diseños evolutivos sean el único modo de investigar dentro de esta disciplina.

Como punto de partida, hay que señalar que la edad no es una variable manipulable. Eso hace que los diseños evolutivos se consideren diseños *ex post facto*. El que sean retrospectivos o prospectivos dependerá de las circunstancias concretas de cada investigación. Pero lo mejor será que nos situemos en un contexto concreto para poder ir desgranando lo que caracteriza a las tres variantes de diseños evolutivos que vamos a presentar.

Aunque es muy habitual situar los diseños evolutivos dentro de un contexto de estudio del desarrollo humano anterior a la edad adulta, esta vez vamos a movernos hacia el final del ciclo vital humano (*life span*).

Cabe considerar que el aumento de la esperanza de vida y los cambios en la pirámide poblacional que muestran un envejecimiento progresivo de la población en los países occidentales han hecho que el interés por la Psicología de la

31. Puede consultar Pérez-Llantada y Fontes (1990).

Vejez haya aumentado en las últimas décadas. Este interés está encaminado a adquirir el conocimiento necesario para potenciar el máximo nivel de calidad de vida a nuestros mayores. En el caso de la Psicología, potenciar al máximo la conservación de las capacidades implicadas en dicha calidad.

Hablamos de conservación porque parece cosa sabida que el envejecimiento implica una pérdida de facultades. Ahora bien, ¿cuántas, cuáles y a qué ritmo se deterioran? Vamos a centrarnos únicamente en el caso de la memoria y, de la mano de Craik, Anderson, Kerr y Li (1995), reflexionaremos sobre las características de los estudios realizados en este campo durante los últimos años³².

3.2. Diseños evolutivos transversales

Convendrá con nosotros que el tema de la memoria es muy amplio y que se hace necesario concretar más. En la revisión que acabamos de citar se hace referencia a casi una docena de "tipos de memoria" o, si lo prefiere, de aspectos diferentes de nuestra capacidad para el recuerdo. Dentro de la distinción ya clásica entre memoria a corto plazo y memoria a largo plazo, vamos a quedarnos sólo con un aspecto de la segunda. Concretamente vamos a proponer que piense la manera de mostrar que las personas de más edad tienen problemas en el proceso de codificación de la información reciente y que estos problemas de codificación impiden incorporar nueva información en el almacén a largo plazo. Éste se mantiene en buen uso por cuanto que nuestros mayores no tienen problemas de acceso a recuerdos de acontecimientos muy lejanos en el tiempo, las famosas "batallas del abuelo".

Podemos entonces establecer que nuestra variable dependiente va a ser el grado de elaboración de la información y la cantidad de recuerdo de los participantes. Una manera sencilla de operativizar esto es presentar un conjunto de frases y pedir que se elabore una explicación de su significado. Unos minutos después se hace una tarea de recuerdo de palabras incluidas en tales frases. Si queremos contrastar la hipótesis de que la edad influye sobre ese grado de elaboración y sobre el recuerdo, un modo de llevar a cabo nuestro estudio es comparar a dos grupos de adultos de diferentes edades en su capacidad de

32. Agradecemos a nuestro compañero Emilio Fernández Lagunilla el que nos facilitara los materiales para la elaboración de este ejemplo.

elaboración y de recuerdo del material elaborado. Podríamos estudiar a un grupo de adultos de cincuenta y cinco años y compararlos con otro de sesenta y cinco.

Eso fue lo que hicieron Hashtroudi, Parker, Luis y Reisen (1989 citado en Craik *et al*, 1995), y encontraron que, efectivamente, los adultos más jóvenes elaboraban más sus explicaciones y de forma más precisa que lo hacían los adultos mayores. Además, la proporción de palabras recordadas también era superior entre los más jóvenes. Su conclusión fue en la línea de apoyar la hipótesis de las diferencias en la riqueza de la elaboración en el momento de la codificación como factor que explica los problemas en el almacenamiento de nueva información que es característico de las personas mayores.

¿Qué pegas se pueden poner a esta conclusión? La mayor amenaza para la validez del estudio vendría de la mano del sesgo de selección. Podría darse el caso de que algunas características de las personas participantes en el estudio se enmascararan con la edad, que no es una variable manipulada, sino, simplemente, seleccionada. Ése podría ser el caso, perfectamente, del nivel de estudios, del grado de actividad intelectual que se realiza habitualmente, de la salud de unos y otros, etc.

El diseño que hemos descrito se denomina *diseño evolutivo transversal* y consiste en comparar en la variable dependiente a diferentes grupos de edad en un mismo momento temporal.

Ya hemos visto cuál puede ser su mayor problema. Para afrontarlo podemos hacer que los miembros de los grupos de edad se parezcan lo más posible en otras variables que pudieran influir sobre la dependiente. Estamos razonando como en los diseños experimentales intersujeto.

Necesitamos que los dos grupos participantes sean parecidos en todo lo que no sea la edad. Seguir esa lógica hasta el final nos lleva a una estrategia intrasujeto. Pues bien, se da el caso de que la edad es la única variable no manipulable que puede estudiarse intrasujeto de forma fácil. Veremos cómo.

3.3. Diseños evolutivos longitudinales

Al final del apartado anterior, acabamos de señalar el rasgo fundamental de este nuevo tipo de diseños evolutivos que presentamos.

En la página web se dispone de una actividad complementaria (capítulo IV, apartado 3.3).

Habría caído en la cuenta de que, después de diez años, las personas de cincuenta y cinco años pasan a tener sesenta y cinco. Sería factible aplicar dos veces el mismo tipo de prueba de memoria a un grupo de adultos con un intervalo de diez años. ¿Qué ganaríamos con ello? Tendríamos mayor seguridad de que las posibles diferencias en el grado de elaboración y recuerdo entre los dos niveles de edad serían debidas a ésta y no a alguna otra variable del estilo de las comentadas anteriormente, siempre y cuando constatáramos que se habían mantenido los hábitos de salud e intelectuales.

Pero esta ganancia de validez interna nos costaría, en este ejemplo, diez años de espera. Eso hace que sea difícil encontrar estudios evolutivos longitudinales en la literatura especializada en el deterioro de la memoria. Como ejemplo reciente puede consultar Rubin *et al* (1998). En esa investigación se hace un seguimiento anual de un grupo de 82 ancianos sanos a lo largo de más de quince años. Se los evalúa clínica y cognitivamente (incluidas pruebas de memoria) y se estudia la relación de algunas variables medidas en años sucesivos con la posible aparición de síntomas de demencia. En cualquier caso, lo que interesa aquí es que tome conciencia de las ventajas que conlleva el diseño longitudinal frente al transversal.

Supongamos que el estudio de Hashtroudi *et al* (1989) hubiera sido de carácter longitudinal, es decir, que las comparaciones de la memoria entre personas de 55 y 65 años se hubieran hecho intrasujeto. Un grupo de adultos que a la sazón –pongamos 1978– tenían cincuenta y cinco años es sometido a la tarea de la elaboración del significado de las frases presentadas. Diez años más tarde, en 1988, se repite la tarea y se presentan los datos, que se publican al año siguiente. Supongamos ahora que en su universidad le hubieran encargado replicar el estudio y hubiera empezado a trabajar en ello en 1991. Habría elegido entonces un grupo de personas adultas de 55 años y habría medido el grado de la elaboración y el recuerdo. Acaba de recoger los datos de la segunda aplicación y ha encontrado que en su estudio no se dan las diferencias que había en el original. ¿Estaría completamente seguro de que, en la población de la que ha extraído la muestra, no se da el deterioro de la memoria observado en el estudio que está replicando?

En nuestra opinión sería más probable sostener la posibilidad de que estuviéramos ante un efecto de cohorte. Si es así, este efecto puede convertirse en una variable enmascarada. Es decir, los cambios que el primer estudio atribuía a la edad podrían deberse a la cohorte. (En los diseños cuasi experimentales hemos explicado el significado de la palabra *cohorte*).

Aunque en ambos estudios nos asegurábamos de que los niveles de estudios, los hábitos de trabajo intelectual y los hábitos de salud se mantenían estables durante los diez años, podría haber ocurrido que las dos cohortes se hubieran comportado de forma diferente respecto a alguno de estos aspectos, de estas variables. Al ser cada vez más durante los últimos años los estudios que trascienden a la población general sobre hábitos de salud, bien podría ser que la generación de nuestro estudio estuviera mejor informada y mantuviera, por tanto, mejores hábitos de actividad intelectual y de salud. ¿Existe algún modo de pasar de la especulación a la comprobación empírica de esta hipótesis? Sí, lo hay: se llama *diseño evolutivo secuencial*.

3.4. Diseños evolutivos secuenciales

Los diseños evolutivos secuenciales, se caracterizan por estudiar longitudinalmente a varias cohortes diferentes.

Suponiendo que Hashtroudi *et al* (1989) se hubieran planteado realizar este tipo de estudio, habrían procedido de la manera siguiente:

- 1) En 1978, aplican la tarea de la elaboración y el recuerdo a un grupo de personas que, a la sazón, tienen 55 años.
- 2) Diez años más tarde vuelven a aplicar la misma tarea al mismo grupo de personas. Además seleccionan a otro grupo que por esas fechas cuenta con 55 años y también les aplican la tarea mencionada.
- 3) Diez años más tarde vuelven a aplicar la tarea al segundo grupo. De ese modo recogen dos aplicaciones para cada cohorte, con una diferencia de diez años entre cohortes y entre aplicaciones.

Quizá la Tabla 4.3 que presentamos a continuación le ayude a verlo más claro. Fíjese bien, porque a continuación le vamos a preguntar alguna cosa al respecto.

Tabla 4.3. Esquema de un diseño secuencial

Cohorte	Fecha de recogida de datos		
	1978	1988	1998
1923	55 años	65 años	
1933		55 años	65 años

En la página web se dispone de una actividad complementaria (capítulo IV, apartado 3.4).

Pensará que hay que estar loco para embarcarse en un proyecto de investigación que dure veinte años. Bueno, no hay estudios sistemáticos que relacionen variables de personalidad con la dedicación a este tipo de trabajos. En relación con el estudio del deterioro de la memoria no conocemos que se hayan llevado a cabo, pero en el terreno del deterioro de habilidades intelectuales hay muchos. El profesor Schaie, por ejemplo, ha sido uno de los autores más prolíficos en este tipo de investigación, tanto que algunos denominan los diseños secuenciales *diseños Schaie*³³.

Dando un repaso final señalaremos varios aspectos que hay que retener:

- El primero es que la edad, cuando se estudia como variable independiente, da lugar a un esquema de investigación *ex post facto*.
- Eso hace que la mayor amenaza de los estudios evolutivos sea la aparición del sesgo de selección. Dicho sesgo es paliable porque la edad es una variable que puede ser estudiada de forma intrasujeto mediante diseños evolutivos longitudinales. Eso hace más difícil la aparición del sesgo de selección.
- Un caso particular del sesgo de selección es el efecto cohorte. Dicho efecto no queda eliminado por los estudios longitudinales pero sí por los diseños evolutivos secuenciales, que lo incluyen como variable independiente añadida a la edad.

4. Resolución de casos prácticos

Presentamos a continuación una serie de resúmenes de investigaciones llevadas a cabo en nuestro país durante la última década, todas ellas publicadas en revistas españolas, mediante los cuales pondremos a prueba su asimilación de los contenidos de este último capítulo del libro. Aunque, en general, hemos sido fieles a las investigaciones originales que reseñamos, se han introducido pequeñas adaptaciones para adecuarlas a los objetivos de esta actividad.

33. Consulte Schaie y Hertzog (1983).

Junto con el resumen, en la página web, le presentamos un conjunto de preguntas con tres alternativas de respuesta de las cuales sólo una es verdadera. Sea cual sea la alternativa que elija, encontrará un comentario que le ayudará a valorar el grado de asimilación de la materia y sus posibles "puntos débiles".

Caso 1: Un programa para mejorar la comprensión lectora.
(Adaptado a partir de Vidal-Abarca (1990)).

Durante las dos últimas décadas, una de las aportaciones más productivas de la Psicología Cognitiva a la enseñanza ha sido la que se ha centrado en estudiar los procesos implicados en la comprensión lectora. Aunque desde el punto de vista del desarrollo teórico quizá hay exceso de modelos y microteorías, desde el punto de vista aplicado se ha puesto de manifiesto que algunas estrategias relativamente sencillas de implantar pueden tener un gran potencial para la mejora de las habilidades de comprensión lectora ya desde la educación primaria. Este trabajo se sitúa dentro de esta línea de investigación y, concretamente, trata de poner a prueba la eficacia de un programa de entrenamiento en la detección de la idea principal explícita de un texto como modo de mejorar su comprensión y su recuerdo.

A 32 niños y niñas de una escuela valenciana se les aplicó un programa compuesto de doce sesiones de 45 minutos repartidas entre los meses de abril y mayo. Otros 32 niños de otra escuela próxima sirvieron como grupo de control. Se tomaron medidas del desarrollo de la lecto-escritura en ambos grupos para constatar que no había diferencias entre los grupos al inicio de la investigación. Se midieron, en ambos grupos, varios aspectos de la comprensión lectora al final de la aplicación del programa. En la Tabla 4.4 se presentan los datos relativos a la detección de la idea principal implícita (la medida podía tomar valores entre 0 y 2) en ambos grupos antes y después de la aplicación del programa.

Tabla 4.4. Detección de ideas principales implícitas antes y después de la aplicación del programa

	Antes	Después	Media
Grupo experimental	0,734	1,063	0,898
Grupo de cuasi control	0,844	0,672	0,758
Media	0,789	0,867	0,828

En la página web se dispone de una actividad complementaria (capítulo IV, apartado 4).

Caso 2: La interacción social en niños andaluces y holandeses.
(Adaptado a partir de Goudena y Sánchez (1996)).

Aunque la interacción social entre iguales es un fenómeno ampliamente estudiado, Goudena y Sánchez constatan que se han realizado pocos trabajos en los que se estudien las variaciones culturales en los procesos de interacción. Y si la escasez de estudios es clara en el caso de las comparaciones de la cultura occidental con otras culturas, más lo es en el caso de la existencia de análisis comparativos dentro de la propia cultura occidental. El objetivo del estudio que llevaron a cabo fue, precisamente, contribuir a llenar ese vacío de trabajos que estudien variaciones en grupos pertenecientes a la cultura occidental.

Ochenta niños y niñas preescolares, de dos grupos de edad (tres y cinco años), fueron observados mientras realizaban juego libre en el patio del recreo mediante un código de observación denominado *Escala de Observación del Juego* (POS son sus siglas en inglés). La mitad de los participantes eran andaluces, veinte del grupo de tres años y veinte del grupo de cinco años. Los cuarenta participantes holandeses se repartían igualmente en dos grupos; veinte del nivel de tres años y veinte del nivel de cinco años. Una de las variables de la comparación fue la implicación de cada niño en juegos de grupo. Se contabilizó el número de intervalos –de un total de treinta– en los que el niño estaba implicado en este tipo de juegos. En la Tabla 4.5 se muestran las medias para los cuatro grupos observados.

Tabla 4.5. Medias de número de implicaciones en juegos de grupo

	Andaluces	Holandeses	Media
Cinco años	24	13	18,5
Tres años	17	14	15,5
Media	20,5	13,5	17

En la página web se dispone de una actividad complementaria (capítulo IV, apartado 4).

Caso 3: Optimismo y pesimismo. Implicaciones para el afrontamiento de dificultades. (Adaptado a partir de Fernández y Bermúdez (2000)).

Desde finales de los años ochenta, dentro del estudio de la personalidad, las estrategias cognitivas de afrontamiento se han considerado como constructos de nivel intermedio entre variables conductuales y variables generales clásicas. Dentro del contexto de logro se han identificado los constructos *pesimismo defensivo* y *optimismo*. Tanto los sujetos pesimistas como los optimistas tienden a rendir de forma similarmente eficaz en tareas que impliquen alcanzar un resultado dentro de un estándar. Los pesimistas defensivos se caracterizan por fijar expectativas bajas aunque tengan una experiencia previa de éxito en una tarea. Ello les lleva a concentrarse en los detalles de la situación para aumentar el control y reducir la ansiedad que les produce su baja expectativa. Los optimistas hacen lo contrario, debido a su experiencia anterior generan altas expectativas y se despreocupan de los detalles particulares, de modo que aparece la ansiedad sólo cuando el resultado se torna negativo.

Los autores de este trabajo constatan la falta de estudios que permitan manipular la experiencia previa en una tarea como modo de contrastar su efecto diferencial en el comportamiento de las personas que utilizan una u otra de las estrategias cognitivas mencionadas. A partir de una muestra inicial de 425 sujetos, estudiantes del primer ciclo de la UNED, se seleccionaron ochenta que se caracterizaban como pesimistas defensivos (cuarenta) y optimistas (cuarenta), asegurándose de que en ambos grupos hubiera el mismo número de varones y mujeres. A ambos grupos se les aplicó una tarea de solución de anagramas en dos condiciones diferentes: dificultad inducida baja y dificultad inducida alta. La inducción consistía en que los cinco primeros anagramas de prueba eran de dificultad baja o alta, y los de la prueba definitiva eran del mismo nivel de dificultad para todos los participantes en todas las condiciones. La mitad de la muestra, al azar, hizo las dos tareas en orden de dificultad inducida baja-alta y la otra mitad en el orden inverso, alta-baja. En la Tabla 4.6 se presentan los resultados relativos al rendimiento de los dos grupos en las dos condiciones.

Tabla 4.6. Rendimiento en la tarea en función de la estrategia y la dificultad percibida

	Dificultad baja	Dificultad alta	Media
Pesimistas	13	13	13
Optimistas	16	10	13
Media	14,5	11,5	13

En la página web se dispone de una actividad complementaria (capítulo IV, apartado 4).

Caso 4: Eficacia del cambio en el estilo atribucional como tratamiento de la depresión leve. (Adaptado a partir de Rodríguez-Naranjo y Godoy (1997)).

El trastorno del estado de ánimo conocido como *depresión* es uno de los que mayor cantidad de atención ha recibido en la literatura psicológica. Tanto desde la visión conductista de la Psicología como desde perspectivas cognitivas, se han aportado diferentes modelos teóricos para explicar la aparición y el mantenimiento del trastorno. Por otro lado, han sido muchos los investigadores que se han preocupado de estudiar la eficacia de diferentes propuestas de trabajo terapéutico para combatir este trastorno. Los autores de este trabajo llaman la atención de la comunidad científica acerca de una relativa descoordinación entre los desarrollos teóricos y las propuestas de intervención terapéutica.

El estudio se centra en el modelo propuesto por Abramson, Seligman y Teasdale (1978). Este modelo postula como causa de la depresión la existencia de un patrón atribucional caracterizado fundamentalmente por el uso de atribuciones internas, permanentes y globales de los resultados negativos. Rodríguez-Naranjo y Godoy aplicaron un entrenamiento atribucional encaminado a cambiar a este patrón, pero lo aplicaron a dos tipos de depresivos: unos que presentaban el patrón atribucional mencionado, al que denominaron *grupo con disforia cognitiva*, y otro que, además, presentaba síntomas de tipo conductual –falta de habilidades sociales–, al que llamaron *grupo con disforia conductual*. A partir de una muestra de 236 adolescentes, seleccionaron mediante un cuestionario de depresión a aquellos que tenían el trastorno. Según sus características atribucionales y de habilidades sociales eligieron dos grupos de treinta que cumplieran las condiciones antes mencionadas. Eliminaron a los individuos que hubieran recibido algún tipo de tratamiento previo.

Las puntuaciones en el cuestionario de depresión, antes y después del tratamiento, se presentan en la Tabla 4.7.

Tabla 4.7.

	Antes	Después	Media
Disforia cognitiva	20	10	15
Disforia conductual	19	19	19
Media	19,5	14,5	17

En la página web se dispone de una actividad complementaria (capítulo IV, apartado 4).

Capítulo V

Anexo

1. Algunas controversias metodológicas

1.1. Las metodologías cualitativas como modo alternativo de investigación

El objetivo de este apéndice es abrir un espacio, aunque sea pequeño, para la reflexión crítica acerca de lo expuesto en la conexión epistemológica. Un modo de ser autocrítico cuando se está bastante convencido de las propias posiciones es dar entrada a las posturas críticas ajenas en una suerte de examen de conciencia dialéctico y desdoblado.

Podría ser que algún conocido o conocida, estudiante de ciencias sociales o educación, al leer el índice de temas de esta obra titulada *Metodologías científicas en Psicología*, torciera el gesto y exclamara algo así como: "¡Esto es metodología cuantitativa!". Usted, que no sabe de que le hablan, piensa que esto es metodología, a secas. Pero, a juzgar por la expresión de disgusto de su interlocutor o interlocutora, sospecha que la cosa no está tan clara y que algo que se le escapa hace que esa persona considere mal la metodología que estudia.

Y es que durante las últimas tres o cuatro décadas, sobre todo en la última, se ha reanudado una discusión que nunca se había zanjado del todo, pero que aparece con mayor virulencia al hilo del surgimiento del movimiento cultural denominado *postmodernidad*. Nosotros vamos a centrarnos en las implicaciones que tiene acusar a una metodología de ser cuantitativa.

La primera implicación, y la más obvia, es que existe otra metodología, "la buena", que en este caso, por contraposición semántica, no puede ser otra que la cualitativa. A nosotros el uso de ambas etiquetas no nos parece muy afortunado³⁴, pero, hoy por hoy, se ha impuesto en nuestro ámbito disciplinar y es necesario utilizarlas. Esperamos que al final entienda nuestro reparo. De mo-

34. Puede consultar Morales y Moreno (1993).

mento le invitamos a que lea con detenimiento el cuadro A.1, en el que presentamos nuestra propia versión de la contraposición entre los dos tipos de metodología³⁵.

Cuadro A.1. Antagonismos entre las metodologías cualitativas y cuantitativas

Metodologías cualitativas	Dimensión comparativa	Metodologías cuantitativas
Subjetivismo, fenomenología	1. <i>Epistemología</i>	Objetivismo, positivismo
Describir desde dentro/ Etnografía Comprender/ Estudio de casos Transformar/ Investigación-Acción	2. <i>Objetivos/Planes de investigación</i>	Describir desde fuera/ Estudios observacionales y de encuestas Clasificar/ Estudios "ex post facto" retrospectivos Predecir/ Estudios "ex post facto" prospectivos Explicar/ Experimentos Intervenir/ Cuasi experimentos
Observación participante Entrevistas no estructuradas Cuestionarios abiertos Narraciones personales Análisis de documentos	3. <i>Técnicas de recogida de información</i>	Observación sistemática Entrevista estructurada Cuestionarios estandarizados Test Tareas estructuradas
Muestras no probabilísticas Análisis cualitativo de datos Rechazo de la estadística	4. <i>Técnicas de selección y análisis de datos</i>	Muestras probabilísticas Análisis cuantitativo de datos Técnicas estadísticas de análisis
Triangulación Reflexión crítica Relevancia subjetiva	5. <i>Criterios de valoración del proceso de investigación</i>	Fiabilidad Validez interna y externa Sensibilidad Generalizabilidad

Las cinco dimensiones propuestas reflejan diferentes niveles de análisis relativos a la metodología.

A continuación iniciamos un breve recorrido por cada una de ellas con objeto de ampliar la información recogida en el cuadro.

1) La primera es la más general y se refiere a la forma en que se concibe la actividad investigadora y su producto: el conocimiento (episteme). En la visión tradicional de la ciencia, en la que la Psicología se entronca desde finales del siglo XIX, el conocimiento que se genera es objetivo. El positivismo se puede aplicar tanto a las

35. Puede consultar Reichardt y Cook (1982); Cohen y Manion (1994).

Ciencias Naturales como a las Ciencias Sociales. Sin embargo, los investigadores críticos con esa visión plantean el hecho de que el conocimiento psicológico siempre será subjetivo si quiere ser relevante para las personas, por lo tanto la visión objetiva es imposible, hay que investigar fenomenológicamente. Es decir, lo que hay que estudiar es lo que las personas perciben como relevante y no imponerles una visión –una teoría– externa a ellos. En el último epígrafe volveremos sobre este asunto.

2) Cuando pasamos a la siguiente dimensión de la comparación nos encontramos con algo bastante más concreto: los planes de investigación y los objetivos a los que responden. Si se fija, las palabras utilizadas para etiquetar las diferentes opciones de los objetivos de investigación de cada una de las dos metodologías contrapuestas se podrían utilizar como sinónimos. Ese es el caso de "transformar" frente a "intervenir", o de "comprender" frente a "explicar". Obviamente el objetivo descriptivo aparece en las dos pero con el énfasis puesto en diferentes puntos de vista. Aunque le parezca que este fenómeno de las etiquetas equivalentes responde a una suerte de pedantería académica le tenemos que advertir que la elección de esos términos tiene una tradición de decenas de años y tratan de poner de manifiesto la diferencia de posturas, aunque para la persona profana resulten indistinguibles. Observe que en la columna de las metodologías cuantitativas aparecen casi todos los nombres de los diferentes tipos de diseños que forman parte de esta obra. Note también que, salvo excepciones –como pueda ser el estudio de casos–, en esta obra no se mencionan ninguno de los planes de investigación que aparecen en la columna de la izquierda. No es una decisión perversa (si lo fuera, no le presentaríamos este apartado). Es una decisión didáctica. Se valora la tradición clásica como mejor punto de anclaje para su formación metodológica. Entendiendo la metodología cuantitativa entenderá la cualitativa. Además al ser la cuantitativa más reglada, facilita el proceso de formación del investigador novato. También, por qué no decirlo, nosotros preferimos investigar mediante metodologías cuantitativas. Al menos así ha sido hasta ahora.

3) El nivel relativo a las técnicas de recogida de información presenta una contraposición casi perfecta entre las dos columnas de la tabla. Para cada técnica de un tipo hay otra simétrica del otro. La contraposición se produce, en este caso, en relación con el grado de estructuración de la técnica. Las técnicas cualitativas, desarrolladas al servicio de una epistemología subjetivista, enfatizan la apertura de la herramienta para hacer máxima la posibilidad de abarcar el fenómeno en toda su riqueza y particularidad. Las técnicas cuantitativas, por el contrario, nacen al servicio de una visión del conocimiento que prima lo común, lo objetivable, lo estructurado, así que la técnica tiene que ayudar en ese proceso de acotación del fenómeno.

4) El cuarto nivel de análisis de las diferencias entre ambos tipos de metodologías hace referencia a las técnicas de selección y análisis de los datos. En este nivel la contraposición es perfecta. Frente a la necesidad de representar la subjetividad la selección de los casos para su estudio no puede ser aleatoria si no todo lo contrario, intencionada, a propósito. Por el contrario, el azar juega un importante papel en la tarea de representar determinados valores –estratos– de determinadas variables que son relevantes para el estudio. Si no los representamos adecuadamente, el estudio estará sesgado, perderá objetividad. Por parecidas razones, los números y –sobre todo– los estadísticos son una herramienta perversa para el objetivo de comprender lo sin-

gular, lo subjetivo, lo fenomenológico. Es necesario mantener el lenguaje natural como código en el que analizar los datos –eso significa la expresión "análisis cualitativo de datos"– porque así cada participante podrá transmitir su propia visión subjetiva sin ser forzado a utilizar las categorías del investigador ni ser diluido en el estadístico que, muchas veces, es un mal resumen de cosas cualitativamente diferentes. Todo lo contrario ocurre en el caso del investigador que trabaja de forma cuantitativa. Para él los números –más allá de la presunción pitagórica de armonía entre naturaleza y matemáticas–, son una herramienta eficaz mediante la que objetivar los fenómenos y la toma de decisiones sobre sus hipótesis. Precisamente porque su mala utilización puede dar lugar a malos entendidos se da mucha importancia a la instrucción estadística del futuro investigador.

5) La última dimensión de comparación es la de los criterios de valoración del proceso investigador. Aunque los conceptos que aparecen a ambos lados de la tabla le pueden parecer todavía desconocidos, no vamos a entrar en detalles con respecto a su significado ya que excede el objetivo de este epígrafe. Lo único que pretendemos al incluirlos es hacer énfasis en que ambos tipos de metodologías tienen sus mecanismos de autorregulación –recuerde que esto era una de las características en la definición de ciencia más laxa que vimos en el primer apartado de este capítulo. Y hacemos énfasis en ello para que tenga algún argumento frente a los que opinan que optar por la metodología cualitativa es escapar a cualquier modo de control, es practicar el anarquismo metodológico. A nosotros nos parece que no es el caso.

Usualmente, el nivel más general –la epistemología– condiciona el proceso de creación en los demás niveles. Eso ha sido históricamente así. Pero una vez creada, pongamos por caso, una determinada técnica de recogida de información, podría darse la situación de que un investigador la utilizara dentro de un contexto epistemológico diferente del contexto en que fue creada³⁶.

Un bonito ejemplo de eso se encuentra en el trabajo de Bem y Allen (1980) cuando utilizan la técnica "Q", subjetivista por excelencia, como modo de abordar el problema del estudio de la interacción entre las dimensiones personales y situacionales a la hora de predecir la conducta.

¿Por qué sacamos a colación este hecho? Porque entre los investigadores de la tradición clásica con mentalidad más abierta han surgido muchas voces en contra de la confrontación entre los dos tipos de metodologías. El caso más conocido quizás sea el de Reichardt y Cook (1986), que en un capítulo titulado enfáticamente "Más allá de los métodos cualitativos versus los cuantitativos" señalan que separando el nivel epistemológico del tecnológico la confrontación desaparece. Este tipo de discursos con voluntad superadora pueden haber dado lugar a investigaciones en las que se utilicen de forma flexible diferentes opciones en los distintos niveles señalados, al menos en

36. Consulte Smith, Harré y Langenhove (1995).

distintos momentos de la investigación. Así, podemos encontrarnos con que la etiqueta *cualitativo* se aplica a una parte del análisis de los datos de un experimento que se han tomado de forma no sistemática y que no se analizan estadísticamente, o al uso de una técnica no estructurada dentro de una fase diagnóstica en una valoración de programas que utiliza un diseño cuasi experimental, o a la fase de elaboración de un código de observación que se va a utilizar en una descripción sistemática de un nuevo problema, etc. Es decir, puede significar cosas muy diversas y resultar perfectamente compatible con la visión metodológica clásica.

Pero esos casos no necesitan más explicaciones. En el apartado siguiente, trataremos de convenceros de que *metodología cualitativa* puede llegar a significar en muchos casos una visión de la investigación totalmente diferente de la que se maneja en esta obra.

1.2. El renacimiento subjetivista

Cuando se cuenta la Historia de la Psicología, se suele recurrir a un relato de su emancipación de la Filosofía para convertirse en ciencia. Y suele mencionarse la creación del laboratorio de Leipzig como el hito clave de dicha emancipación. Sin embargo, este relato es muy opaco a algunos problemas que preocupaban a los psicólogos de la época. El propio Wundt, fundador del mencionado laboratorio, defendía la necesidad de mantener dos tipos de disciplinas en el conocimiento humano: las ciencias naturales y las ciencias del espíritu. Parte de la Psicología debía, en su opinión, ubicarse en este último grupo. Y es que la adscripción a la visión objetivista de las Ciencias Naturales genera una paradoja en la Psicología, ya que "ha soñado la compleja intención de establecer un conocimiento objetivo acerca de las raíces mismas de la subjetividad humana" (Rivière, 1991, p. 168).

Las pasadas décadas de los ochenta y los noventa han representado una recuperación de la sensibilidad subjetivista en todos los ámbitos de las Ciencias Sociales, la Psicología y la Educación, y dicha recuperación viene de la mano de dos tipos de críticas a las ciencias oficiales: una crítica epistemológica y otra social.

La primera está contenida en la visión postmoderna de la cultura. Frente al proyecto renacentista de alcanzar la verdad –una única verdad– en todos los órdenes del conocimiento mediante el uso de la razón –y, por tanto, la supera-

ción de la visión teológica medieval de la filosofía– se contraponen la actitud de tolerancia entre teorías, el relativismo epistemológico. Frente al paradigma dominante, disciplinas pluriparadigmáticas.

La otra crítica, muy asociada a la anterior, viene a reprochar que, en muchos casos, la verdad de la Ciencia se ha puesto al servicio del orden establecido. Por tanto, se reclama para el científico social, para el educador, para el psicólogo, un papel crítico y no legitimador de un supuesto pensamiento único defendido también desde las instituciones científicas. De ese modo los criterios de valoración ya no ponen hincapié en la fiabilidad o la validez interna de la investigación, sino en su relevancia social, su compromiso, su capacidad de ser vehículo de expresión de subjetividades ajenas³⁷.

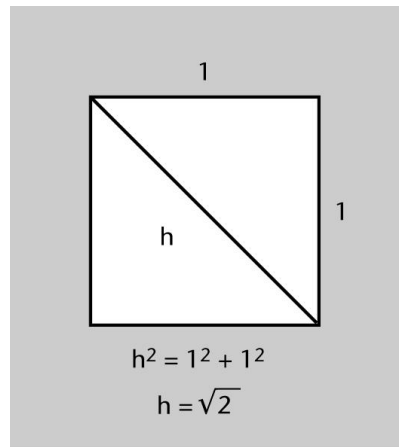
¿Por qué hacemos el esfuerzo de presentarle una visión tan alejada de la nuestra? Porque está ahí y porque si en su futuro ejercicio profesional se sitúa como "científico" es probable que sea criticado por postular sus argumentos como "objetivos".

¿Es resoluble el conflicto? Nuestra opinión al respecto es que se hace necesario recurrir al concepto de inconmensurabilidad entre paradigmas acuñado por Khun (1971). Dicho concepto establece que las disputas entre diferentes paradigmas –entre distintos modos de entender los conocimientos y el modo de trabajar dentro de un ámbito disciplinar– no es resoluble mediante el recurso a una métrica común que nos permita tomar la decisión acerca de cuál es el mejor.

Aunque a las personas con formación matemática ese concepto de inconmensurabilidad les parece típico de filósofos y, por tanto, sospechoso de vacuidad y mera retórica, hay que señalar algo que no suele hacerse. A saber, los primeros en utilizar el concepto fueron los antiguos matemáticos pitagóricos para expresar las relaciones entre la diagonal y el lado de un cuadrado, cuando este último vale la unidad. En ese caso la diagonal es una hipotenusa de un triángulo isósceles con dos catetos de valor 1. El teorema de Pitágoras establece que el valor de dicha hipotenusa será la raíz cuadrada de 2. Los pitagóricos no conocían ningún número cuyo cuadrado fuera 2, así que decidieron llamar a tal número *alogon*, 'irracional'³⁸.

37. Puede consultar Denzin y Lincoln (1998); Smith, Harré y Langenhove (1995).

38. Véase Guedj (2000).



Siendo inconmensurables los dos paradigmas metodológicos, al tratar de medir desde el conocido –el cuantitativo– las dimensiones del desconocido –el cualitativo– la primera reacción es calificarlo de irracional. Esperamos que en su caso no se repita la historia.

Una sugerencia, como comentario final a este apartado de las controversias metodológicas. Creemos que le resultará mucho más interesante y comprensible cuando lo vuelva a leer una vez terminada toda la obra.

2. La estructura del informe científico: las normas de publicación APA '01

Una buena investigación no publicada es como una partitura genial metida en un cajón. Necesita ser difundida para que pueda hacerse uso de las aportaciones propuestas. Los resultados de las investigaciones se publican en revistas especializadas y deben ajustarse a un formato determinado. Esta forma no es optativa para los investigadores, es una especie de protocolo aceptado para que la comunicación sea más fluida. No hay lugar para la originalidad en la estructura de la comunicación científica en Psicología. En los trabajos empíricos la estructura es ésta:

- Título
- Resumen
- Introducción
- Método

Resultados

Discusión

Referencias

Casi todas las revistas de Psicología siguen esta estructura. En una de las páginas finales tienen un resumen de las normas de publicación, con detalles concretos. En estas normas suelen hacer referencia al manual de publicación de la APA (American Psychological Association, 2001), publicado por la Asociación Americana de Psicología³⁹. En este libro (de 439 páginas) no se deja nada al azar, y pueden encontrarse desde consejos para la redacción del título del trabajo hasta el tamaño de letra en el procesador de textos. Lo que puede producir una sensación de encorsetamiento al autor suele redundar en facilidades para el lector. Por ejemplo, la revista española *Psicothema*, en sus normas para la aceptación de trabajos, especifica lo siguiente:

"1. *Psicothema* publica trabajos de carácter empírico que estén hechos con rigor metodológico y que representen una contribución al progreso de cualquier ámbito de la Psicología Científica. También se acogen trabajos de naturaleza teórica [...], con preferencia para aquellos que se comprometan con cuestiones críticas de la investigación y discutan planteamientos polémicos". Y más adelante, "4. La preparación de los manuscritos debe atenerse, por lo demás, a las normas de publicación de la APA."

A continuación puede observar cómo es la estructura de un informe científico teniéndolo a la vista. Cuando lo desee, puede hacer clic sobre la actividad apéndice B 1, de la página web.

Si quiere acceder a otro artículo de investigación elegido por usted, le invitamos a ejecutar la actividad apéndice B 2, de la página web.

A continuación ampliamos las características que definen cada una de las secciones del informe de investigación en Psicología.

En el *resumen* (de 100 a 120 palabras) tiene que darse información, en un solo párrafo, de los aspectos siguientes:

- a) El problema que se ha investigado.
- b) El método (incluyendo aparatos, modo de obtener los datos y quiénes fueron los participantes).
- c) Los hallazgos, las conclusiones y consecuencias de estos hallazgos.

39. <http://www.apa.org/>.

Como puede ver, todo un reto a la capacidad para resumir. Es muy importante que se haga bien, porque ésta será la carta de presentación y la que usarán los posibles lectores para acabar leyendo la investigación.

En la *introducción* –que no lleva título de introducción, sino el título del trabajo– los investigadores tienen que exponer de manera concisa y efectiva los aspectos siguientes:

a) Presentación del problema. Eso debe incluir una justificación de por qué tenía que hacerse esta investigación. Los motivos que habitualmente se esgrimen son: falta de evidencia empírica, datos inconsistentes, resultados contradictorios, posibles explicaciones alternativas a investigaciones ya realizadas, etc. (véase el capítulo 2 de León y Montero (2003)).

b) Presentación de los antecedentes. Sin remontarse a los fenicios, deben exponerse brevemente los trabajos publicados que conducen a esta investigación.

c) Definición del propósito del estudio. Deducción lógica de lo que se pretende hacer y de lo que se espera encontrar.

En la sección del *método* es en la que se proporciona la información para que el trabajo cumpla con el requisito de la replicabilidad, es decir, que otros puedan repetir exactamente la investigación y, presumiblemente, llegar a los mismos resultados. Esta sección tiene las subsecciones siguientes:

a) *Participantes*. Este dato es muy importante, porque permite evaluar el grado de generalización, la comparación con otros grupos o evaluar las diferencias culturales. Tiene que informarse de cómo accedieron a la investigación, si se les pagó, la distribución por género y edad. En general, características demográficas *relevantes* para la investigación.

b) *Aparatos o materiales*. Si alguien quiere repetir la investigación, deberá conocer todos los detalles necesarios para usar las mismas herramientas.

c) *Diseño y procedimiento*. El diseño se refiere al *plan seguido*, las variables independientes, sus niveles, el modo de producirlas (o seleccionarlás) y medirlas, y también, si se han aplicado de forma intersujeto o intrasujeto, las variables dependientes y el modo de medirlas. En cuanto al procedimiento, podemos decir que es el equivalente a las instrucciones de uso. Tiene que ser secuencial, claro y suficiente para repetir todos los pasos de la investigación. Es crucial que se informe de cómo se asignaron los participantes a las condiciones, qué instrucciones –literales– se les dieron y cómo se registraron las medidas.

En la sección de *resultados*, el lector espera encontrar un resumen de los datos registrados y una evaluación de su alcance (significación estadística, tamaño del efecto).

Suele mejorar la comunicabilidad de la presentación si se resumen los datos en tablas y, a veces, si vienen acompañadas de figuras. No obstante, como las revistas limitan el número total de páginas de los manuscritos, los investigadores tienen que evaluar la conveniencia de incluir estos elementos. (Esta sección resultará espesa al lector que no tenga una formación estadística adecuada.)

En la sección de la *discusión*, los investigadores tienen que seguir los pasos siguientes:

a) Empiezan con una evaluación de las hipótesis que plantearon en la introducción: se rechazan o se mantienen, provisionalmente, con los datos encontrados.

b) Se les obliga a que comparen sus hallazgos con las investigaciones publicadas más parecidas.

c) Suelen acabar con las implicaciones teóricas y prácticas de los hallazgos, y también con la orientación para futuros trabajos.

En la sección de *referencias*, aparecen las fuentes documentales citadas en el texto. Las normas para esta sección son muy rígidas y detalladas para cada casuística.

Observe que, además de la información primordial de los autores, título y lugar de publicación, podemos saber si se trata de un libro (León y Montero, 2003), de un capítulo de un libro (Toro y Castro, 2000) o de un artículo en una revista (Godden y Baddeley, 1975). (Consulte estas citas en la sección de referencias de este libro y observe las diferencias entre los tres.)

El propósito de este apéndice no es darle un minicursillo acelerado para que el mes que viene publique algo en una revista, sino familiarizarle con la estructura de la comunicación científica y facilitarle su lectura. Cuando esté en situación de querer publicar algo, le aconsejamos que empiece con los modelos que presentan los textos de metodología y que después pase al propio manual de la APA.

3. La ética en la investigación. El código deontológico del psicólogo

La sociedad que sostiene económicamente a los investigadores espera que el conocimiento que adquieren sea científico. Por científico –diremos sucintamente– entendemos que ofrezca garantías. Para alcanzar esta loable finalidad, los investigadores utilizan metodologías que les garantizan la científicidad del conocimiento adquirido. Algunas de estas metodologías, en el campo de la Psicología, son las que se han expuesto en este libro.

Para aplicar las metodologías expuestas necesitamos el concurso de personas que estén dispuestas a colaborar. En algunos casos, como en el de la observación natural, los participantes apenas son interferidos por los investigadores.

En otros, su vida se ve alterada, como informa la noticia que apareció en los medios (*El País*, 23-03-01, p. 64) acerca de que 27 monjas de conventos de Badajoz se ofrecieron a consumir jamón ibérico durante dieciséis semanas. Con el resto de las variables constantes –y suponemos que con otros conventos de cuasi control–, los investigadores del Hospital de San Sebastián (Badajoz) constataron un descenso en los niveles de colesterol de las "sufridas" participantes.

Quizá también se hubiera apuntado a este cuasi experimento, pero quizá no a los experimentos de Milgran (1974), donde los participantes eran inducidos a creer que aplicaban descargas eléctricas peligrosas a otros seres humanos.

¿Podemos hacer cualquier cosa para conseguir la buena finalidad del conocimiento científico? ¿Es aséptica la metodología respecto a la moral? Toda conducta humana satura en más o menos grado la ética. La del científico también.

Permítanos que citemos algo que no se avala con datos publicados en revistas científicas, sino en una vida comprometida. Gandhi, a lo largo de su vida pública, insistía en que

"las finalidades tienen que estar en los medios, como el árbol en la semilla".

Es la afirmación antitética a la, demasiadas veces esgrimida, de que la "la finalidad justifica los medios". Si seguimos el principio ético de Gandhi, la buena finalidad de la investigación de obtener conocimiento beneficioso no puede justificar medios perjudiciales para obtenerlos. Dicho de otra manera: la investigación tiene que salvaguardar por encima de todo los derechos de los participantes.

No tiene que infligirles ningún tipo de daño permanente o pasajero. Estas consideraciones están recogidas en los códigos éticos elaborados en cada país. En España están recogidas por el Colegio de Psicólogos.

El Colegio Oficial de Psicólogos⁴⁰ es una entidad que agrupa a los profesionales de la Psicología en España. Entre otras iniciativas tiene la de haber elaborado un código deontológico de la profesión. El código deontológico, lógicamente supeditado a las leyes nacionales y de la Unión Europea, orienta la honradez de los psicólogos en todos los ámbitos de su profesión. Este código también sirve para evaluar la conducta de sus asociados, cuando fuera pertinente. En el preámbulo de este código se dice lo siguiente:

"La profesión de psicólogo/a se rige por principios comunes a toda deontología profesional: respeto a la persona, protección de los derechos humanos, sentido de responsabilidad, honestidad, sinceridad para con los clientes, prudencia en la aplicación de instrumentos y técnicas, competencia profesional, *solidez de la fundamentación objetiva y científica de sus intervenciones profesionales*⁴¹."

Como puede verse, una de las exigencias de la práctica profesional es que los procedimientos utilizados se hayan generado de acuerdo con las exigencias científicas que haya aprendido en este libro. Sin embargo, hay otra área que incide de lleno en la actividad investigadora: el título IV. Puede consultarse íntegro en el cuadro apéndice C1 de la página web.

En el título IV se habla de que:

- En la investigación no se puede producir ningún tipo de daño a los participantes y que su participación debe ser consentida de manera explícita.
- Asimismo, cuando se trate de participantes menores de edad, se necesitará el consentimiento explícito de los tutores.
- Aunque se haya firmado la aceptación de las condiciones de la prueba, los participantes deberán saber que están en su derecho de abandonarla en cualquier momento que lo deseen, sin necesidad de justificarlo.
- Si se necesitara presentar la investigación con una apariencia falsa para conseguir sus finalidades, al final de las pruebas el investigador explicará a los participantes el porqué de esta circunstancia, así como los verdaderos objetivos.
- Asimismo, se recogen las circunstancias bajo las cuales es ético infringir algún tipo de daño temporal o permanente cuando se trabaja con animales.

40. Le recomendamos que en algún momento visite su portal en <http://www.cop.es/>.

41. La cursiva del final es nuestra.

Bibliografía

Abramson, L.Y., Seligman, M.E.P., Teasdale, J. D. (1978). Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology*, 87, 49-74.

Alas, L. (1886). *La Regenta*. Madrid: Alianza, 1966.

American Psychological Association (1998). *Manual de estilo de publicaciones de la American Psychological Association* (2ª ed.) Washington: APA, 1994.

Anderson, R. (1980). *Cómo rejuvenecer el cuerpo estirándose*. Barcelona: Integral.

Arkin, H. y Colton, R. (1962). *Tables for statisticians*. Nueva York: Barnes and Noble.

Arnau, J., Anguera, T. y Gómez, J. (Ed.) (1990). *Metodología de la investigación en ciencias del comportamiento*. Murcia: Universidad de Murcia.

Bakeman, R. y Gottman, J.M. (1989). *Observación de la interacción: Introducción al análisis secuencial*. Madrid: Morata, 1986.

Barlow, D.H. y Hersen, M. (1984). *Single case experimental designs* (2ª ed.). Nueva York: Pergamon Press.

Bem, D.J. y Allen, A. (1980). Cómo predecir a algunas personas en algunas situaciones: la búsqueda de consistencias transituacionales en la conducta. *Estudios de Psicología* (ed. original de 1974), 3, 57-74).

Bernard, M.E., Kratochwill, T.R. y Keefauver, L.W. (1983). The effects of rational-emotive therapy and self-instructional training on chronic hair pulling. *Cognitive Therapy and Research*, 7, 273-280.

Berk, L. E. (1986). Relationship of elementary school children's private speech to behavioral accompaniment to task, attention, and task performance. *Developmental Psychology*, 22, 671-680.

Bowen, J. (1990). *Historia de la educación occidental*. Barcelona: Herder.

Breuer, J. y Freud, S. (1893). Sobre los mecanismos físicos de los fenómenos histéricos. *Obras completas de S. Freud* (traducción de J.L. Etcheverry, vol. 2). Buenos Aires: Amorrortu, 1976.

Broncano, F. y Mundó, J. (2001). Filosofía i epistemologia de les ciències particulars. En: J. Mundó (coord.), *Filosofia i epistemologia de la ciencia*. Barcelona: Universitat Oberta de Catalunya.

Campbell, D.T., Stanley, J.C. (1966). *Experimental and quasi-experimental design for research*. Chicago: Rand McNally.

Campos, M., Díez, J.A., Domènech, A., Estany, A., Mundó, J. y Pérez, E. (2001). "Filosofía de la ciencia". En J. Mundó (coord.), *Filosofía i epistemologia de la ciència*. Barcelona: Universitat Oberta de Catalunya.

Chalmers, A.F. (1984). *¿Qué es esa cosa llamada ciencia?* Madrid: Siglo XXI, 1982.

Charcot, M. (1887). *Oeuvres complètes, tome III. Leçons sur les maladies du système nerveux*. París: A. Delahaye et E. Lecrosnier libraires-éditeurs.

Chorot, P. y Martínez-Narváez, P. (1995). Trastornos somatoformes. En: A. Belloch, B. Sandín y F. Ramos (ed.), *Manual de Psicopatología*, 225-270. Madrid: McGraw-Hill.

Chudy, J. F., Jones, G. E. y Dickson, A. L. (1983). *Modified desensitization approach for the treatment of phobic behavior in children: A quasi-experimental case study*. *Journal of Clinical Child Psychology*, 12, 198-201.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20, 37-46.

Cohen, L. y Manion, L. (1994). *Research methods in education* (4ª ed.). Londres: Routledge.

Cone, J.D. (1978). The Behavioral Assessment Grid (BAG): A conceptual framework for behavioral assessment. *Behavior Therapy*, 9, 882-888.

Cook, T.D. y Campbell, D.T. (1979). *Quasi-experimentation design and analysis issues for the field settings*. Boston: Houghton Mifflin.

Craik, F.I.M. y Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology, General*, 104, 268-294.

Craik, F.I.M., Anderson, N.D., Kerr, S.A. y Li, L.Z.H. (1995). Memory changing in normal aging. En A.D. Baddeley, B.A. Wilson y F.N. Watts (Ed.). *Handbook of memory disorders* (211-242). Nueva York: John Wiley and sons.

Dampier, W.C. (1972). *Historia de la ciencia y sus relaciones con la filosofía y la religión*. Madrid: Tecnos.

Darley, J. y Bateson, C. D. (1973). From Jerusalem to Jericho: A study of situational and dispositional variables in helping behavior. *Journal of personality and Social Psychology*, 27, 100-108.

Defensor del Pueblo. (2000). *Violencia escolar: el maltrato entre iguales en la educación secundaria*. Madrid: Informes, estudios y documentos del Defensor del Pueblo.

Denzin, N.K. y Lincoln, Y. (1998). Introduction. Entering the field of qualitative research. En N.K. Denzin y Y. Lincoln (Ed.), *The landscape of qualitative research. Theories and Issues* (pp. 1-34). Thousand Oaks, CA: Sage.

Díaz, R.M. y Berk, L. (ed.) (1992). *Private Speech: From social interaction to self-regulation*. Nueva York: LEA.

- Diges, M.** (1997). *Los falsos recuerdos*. Barcelona: Paidós.
- Durheim, E.** (1897). *El suicidio* (ed. de L. Díaz). Madrid: Akal.
- Echevarría, J.** (1999). *Introducción a la metodología de la ciencia. La filosofía de la ciencia en el siglo XX*. Madrid: Cátedra.
- Eckman, P., Levenson, R.W. y Friesen, W.V.** (1983). Autonomic nervous system activity distinguishes among emotions. *Science* 221, 1208-1210.
- Español, S.** (2000). *Un estudio de semiosis evolutiva* (tesis doctoral no publicada). Madrid: Universidad Autónoma de Madrid.
- Español, S. y Rivièrre, A.** (2000). Gestos comunicativos y contextos interpersonales: un estudio con niños de 10 a 16 meses. *Estudios de Psicología* 65-66, 225-245.
- Eysenk, H.J. y Eysenk, M.W.** (1985). *Personality and individual differences*. Nueva York: Plenum Press.
- Fernández, E. y Bermúdez, J.** (2000). Control situacional y focalización en la tarea en pesimistas defensivos y optimistas. *Estudios de Psicología*, 23-40.
- Frank de, M.G. y Gilovich, T.** (1988). The dark side of self and social perception: Black uniforms and aggression in professional sports. *Journal of Personality and Social Psychology*, 54, 74-85.
- García Bajos, E. y Migueles, M.** (1999). Memoria de testigos en una situación emocional vs. neutra. *Psicológica*, 20, 91-102.
- Goddén, D.R. y Baddeley, A.D.** (1975). Context dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, 66, 325-331.
- Goudena, P.P. y Sánchez, J.A.** (1996). Interacción entre iguales en Andalucía y Holanda. *Infancia y Aprendizaje*, 75, 49-58.
- Guedj, D.** (2000). *El teorema del loro*. Barcelona: Anagrama.
- Hartmann, D.P. y Hall, R.V.** (1976). The changing criterion design. *Journal of Applied Behavior Analysis*, 9, 527-532.
- Hashtroudi, S., Parker, E.S., Luis, J.D. y Reisen, C.A.** (1989). Generation and elaboration in older adults. *Experimental Aging Research*, 15, 73-78.
- Hernández, J.M.** (2000). *La personalidad. Elementos para su estudio*. Madrid: Biblioteca Nueva.
- James, J.** (1890). *The principles of Psychology*. Nueva York: Holt.
- Jones, B. y Kenward, M.** (1989). *Design and analysis of cross-over trials*. Londres: Chapman and Hall Ltd.
- Kahneman, D., Frederickson, B.L., Schreiber, C.A. y Redelmeir, D.A.** (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, 4, 401-405.

Kazdin, A.E. (1982). *Single-case research designs: Methods for clinical and applied settings*. Nueva York: Oxford University Press.

Kerlinger, F. N. y Lee, H.B. (2000). *Foundations of behavioral research* (4ª ed.). Fort Worth, TX: Harcourt College Publishers.

Khun, T.S. (1971). *La estructura de las revoluciones científicas*. Madrid: Fondo de Cultura Económica, 1962.

Latané, B. y Darley, J.M. (1968). Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology*, 215-221.

León, O.G. y Montero, I. (2001). Cómo explicar el concepto de interacción sin estadística: análisis gráfico de todos los casos posibles en un diseño 2x2. *Psicothema*, 13, 159-165.

León, O.G. y Montero, I. (2003). *Diseño de investigaciones. Introducción a la lógica de la investigación en Psicología y Educación* (2ª ed.). Madrid: McGraw-Hill.

Lepper, M.R., Greene, D. y Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the overjustification hypothesis. *Journal of personality and Social Psychology*, 28, 129-137.

Loftus, E.F., Miller, D.G. y Burns, H.J. (1978). Semantic integration of verbal information into visual memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, 4, 19-31.

Maldonado, F.C.R. (1966). *Refranero clásico español* (2ª ed.). Madrid: Taurus.

Marañón, G. (1985). Contribución al estudio de la acción emotiva de la adrenalina. *Estudios de Psicología* (ed. original de 1924), 21, 75-89.

McCann, I. L. y Holmes, D.S. (1984). Influence of aerobic exercise on depression. *Journal of Personality and Social Psychology*, 46, 1142-1147.

Milgram, S. (1974). *Obedience to authority*. Nueva York: Harper.

Montero, I., Pérez del Palacio, M. (En preparación, a). *El desarrollo del habla privada en el contexto del aula de educación infantil*. Manuscrito en preparación.

Montero, I., Pérez del Palacio, M. (En preparación, b). *La influencia del habla privada sobre la resolución de la tarea en situaciones estructuradas*. Manuscrito en preparación.

Morales, F. (Ed.) (1989). *Metodología y Teoría de la Psicología* (5ª ed.). Madrid: UNED.

Morales, M. y Moreno, R. (1993). Problemas en el uso de los términos cualitativo/cuantitativo en Investigación Educativa. *Investigación en la escuela*, 21, 38-50.

Mundó, J. (2001). Introducció a la teoria del coneixement. En J. Mundó (dir.). *Filosofia i epistemologia de la ciència*. Barcelona: Universitat Oberta de Catalunya.

Pérez-Llantada, C. y Fontes, S. (1990). Metodologías de la investigación en Psicología evolutiva. En J.A. García Madruga y P. Lacasa (dir.). *Psicología evolutiva* (pp.181-211). Madrid: UNED.

Pintrich, P.R. y Shunk, D.S. (1996). *Motivation in education. Theory, research and applications*. Englewood Cliffs, NJ: Prentice-Hall.

Popper, K. (1959). *The logic of scientific discovery*. Londres: Hutchinson.

Pulido, A. (1971). *Estadística y técnicas de investigación social*. Madrid: Anaya.

Reichardt, C.S. y Cook, T.D. (1986). Más allá de los métodos cualitativos versus los cuantitativos. En T.D. Cook y C.S. Reichardt (Ed.). *Los métodos cualitativos y cuantitativos en investigación evaluativa*. Madrid: Morata, 1986.

Rivière, A. (1991). *Objetos con mente*. Madrid: Alianza.

Rodríguez-Naranjo, C. y Godoy, A. (1997). Factores que predicen la eficacia de diferentes tratamientos para la depresión leve o disforia. *Psicothema*, 9, 105-117.

Rosa, A. (2000). ¿Qué añade a la Psicología el adjetivo cultural?. *Anuario de Psicología*, 31, 27-57.

Rosnow, R.L. y Rosenthal, R. (1996). *Beginning behavioral research. A conceptual primer* (2ª ed.). Englewood Cliffs, NJ: Prentice-Hall.

Rowling, J. (1999). *Harry Potter y la piedra filosofal*. Barcelona: Emecé, 1997.

Rubin, E., Storandt, M., Miller, P., Kinscherl, D.A., Grant, E.A., Morris, J.C. y Berg, L. (1998). A prospective study of cognitive function and onset of dementia in cognitively healthy elders. *Archives of Neurology*, 395-401.

Rubio, G., Montero, I., Jáuregui, J., Marín, J.J. y Santo-Domingo, J. (1998). Relación entre conducta violenta y tentativas autolíticas. *Archivos de Neurobiología*, 61, 317-326.

Schachter, S. y Singer, J.E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 69, 379-399.

Schaie, K.W. y Hertzog, C. (1983). Fourteen-year cohort-sequential analysis of adult intellectual development. *Developmental Psychology*, 531-543.

Shaughnessy, J.J., Zeichmeister, E.B. y Zeichmeister, J.S. (2003). *Research methods in Psychology* (5ª ed.). Nueva York: McGraw-Hill.

Simpson, F. (1964). *The view of life*. Nueva York: Harcourt, Brace and World.

Smith, J.A., Harré, R. y Van Langenhove, L. (1996). *Rethinking methods in Psychology*. Thousand Oaks, CA: Sage.

Toro, J. y Castro, J. (2000). Trastornos del comportamiento alimentario. En A. Bulbena; G.E. Berrios y P. Fernández de Larrinoa (Ed.). *Medición clínica en psiquiatría y psicología* (pp. 215-224). Barcelona: Masson.

Turnbull, O.H., Stein, L. y Lucas, M.D. (1995). Lateral preferences in adult embrancing: A test of the 'hemispheric asymmetry' theory of infant cradling. *The Journal of Genetic Psychology*, 156, 17-21.

Vidal-Abarca, E. (1990). Un programa para la enseñanza de la comprensión de ideas principales en textos expositivos. *Infancia y Aprendizaje*, 53-72.

Wason, P.C. (1960). On the failure to eliminate hypotheses in a conceptual tasks. *Quarterly Journal of Experimental Psychology*, 12, 129-140.

Wincze, J.P., Leitenberg, H. y Agras, W.S. (1972). The effects of token reinforcement and feedback on the delusional verbal behavior of chronic paranoid schizophrenics. *Journal of Applied Behavior Analysis*, 5, 247-262.

Glosario

acuerdo entre observadores Grado en el que dos observadores registran las mismas categorías cuando observan simultáneamente el mismo fenómeno.

adaptación a las pruebas Amenaza potencial para la validez interna que aparece en los diseños que necesitan repetir la medida de la variable dependiente a los mismos sujetos, ya que éstos pueden modificar su rendimiento por la reiteración de la medida.

afijación proporcional Procedimiento para determinar el número de unidades muestrales de un estrato que debe formar parte de la muestra; se calcula aplicando a la muestra la proporción que el estrato representa en la población.

afijación simple Procedimiento para determinar el número de unidades muestrales de un estrato que debe formar parte de la muestra; se calcula dividiendo el tamaño de la muestra entre el número de estratos en que se divide la población.

aprendizaje Cambio permanente del comportamiento producido en una condición experimental que afecta a la respuesta a las siguientes condiciones. Antes de empezar, hay que asegurarse de que no haya un aprendizaje que invalide el experimento intrasujeto. Lo anterior se exceptúa cuando se estudia el aprendizaje como variable dependiente.

asignación aleatoria Procedimiento basado en el azar para decidir a qué condición experimental va cada participante. Asegura que la probabilidad de todos y cada uno de los participantes para ir a una condición sea la misma.

ciencia Forma de conocimiento que se caracteriza por su objeto –la naturaleza y la sociedad– y su método –el método hipotético-deductivo–.

código arbitrario Sistema de categorías arbitrariamente creadas por el investigador para registrar los elementos del fenómeno de su interés.

conclusión Parte final del proceso de investigación mediante la cual el investigador conecta los datos obtenidos con el objetivo de su estudio.

condición Cada una de las tareas en que se producen los valores de la variable independiente, cuyos efectos sobre la variable dependiente se desean probar.

condición experimental véase condición.

confianza Grado de seguridad estadística asociado al proceso de estimar los valores de la población a partir de los valores descritos en la muestra.

conglomerado Conjunto natural de unidades muestrales. En el proceso de muestreo, se seleccionan los conglomerados en vez de las unidades, como forma de simplificar el proceso.

constructo Variable que se postula teóricamente pero que no es accesible a la observación.

control del orden por reequilibrado Técnica de control del efecto de la práctica en el contexto de los experimentos intrasujeto con dos niveles. Consiste en empezar con una ordenación aleatoria de las condiciones y seguirla por la ordenación inversa. Puede hacerse completo o incompleto.

cuadrado latino Técnica de control del orden de presentación en los diseños intrasujeto. Se caracteriza porque la variable independiente tiene tres o más niveles, y porque cada nivel se presenta una vez en cada uno de los órdenes posibles. Se puede medir el efecto debido al orden de presentación.

cuestionario Conjunto de preguntas (y respuestas posibles) escritas para estudiar en una muestra las características definidas por el objetivo de una investigación por encuesta.

dato Elemento de información en el que se codifican los resultados obtenidos en la investigación y por cuyo análisis se establece la conclusión.

decisión estadística Acto de aceptar o rechazar la hipótesis nula a partir de la probabilidad obtenida para el estadístico utilizado en el contraste.

diseño Plan de investigación cuyas variantes dependen tanto del objetivo de la investigación como de las condiciones en que se lleva a cabo.

diseño AB Tipo de metodología experimental con un solo sujeto caracterizado por una fase de línea base al principio, seguido de una fase de tratamiento. Se utiliza cuando no es conveniente retirar el tratamiento, o cuando tras el tratamiento se producen cambios irreversibles.

diseño ABAB Tipo de diseño experimental con un solo sujeto caracterizado por una fase de línea base al principio, seguido de una fase de tratamiento tras la cual se retira la intervención. Se termina con una nueva fase de tratamiento.

diseño cuasi experimental Diseño que contrasta una relación de causa efecto pero en el cual las circunstancias de su implantación no permiten, *a priori*, establecer los controles mínimos.

diseño cuasi experimental, con cuasi control en una segunda variable dependiente Diseño cuasi experimental en el que el impacto del tratamiento se establece comparando, dentro del mismo grupo, las medidas de la variable dependiente tratada con las de otra variable dependiente no tratada, pero igualmente sensible al tratamiento.

diseño cuasi experimental *pre post*, con cuasi control en una cohorte anterior Diseño cuasi experimental con medidas antes y después de la aplicación del tratamiento en el que el grupo de cuasi control se establece en una cohorte anterior perteneciente a la misma institución.

diseño cuasi experimental *pre post*, con grupo de cuasi control Diseño cuasi experimental en el que el impacto de un tratamiento se establece por la comparación de una medida anterior y otra posterior al tratamiento, tomadas en dos grupos. A uno de ellos se le aplica el tratamiento y al otro no. Los grupos no pueden ser formados aleatoriamente.

diseño cuasi experimental *pre post*, con un solo grupo Diseño cuasi experimental en el que el impacto de un tratamiento se establece mediante la comparación entre una medida anterior y otra posterior a la aplicación del mismo a un único grupo de sujetos.

diseño de cohortes Plan de investigación, en el contexto de la metodología con encuestas, que consiste en estudiar de forma comparativa grupos poblacionales llamados *cohortes*; éstas se definen como conjuntos agrupados por una circunstancia vital, como la fecha de nacimiento, de ingreso en una institución, etc.

diseño de grupos aleatorios Tipo de metodología experimental con distintos sujetos (intersujetos). Se caracteriza por una asignación aleatoria de los participantes a las distintas condiciones experimentales.

diseño de serie temporal interrumpida, con un solo grupo Diseño cuasi experimental en el que el impacto de un tratamiento se establece en un solo grupo mediante la comparación de una serie de medidas anteriores con otra serie de medidas posteriores a la aplicación del tratamiento.

diseño descriptivo Plan de investigación cuyo objetivo es realizar descripciones precisas que permitan establecer regularidades susceptibles de ser postuladas como leyes.

diseño evolutivo Diseño que estudia la edad como variable independiente. Es de tipo *ex post facto*, ya que esta variable no puede ser manipulada por el investigador.

diseño evolutivo longitudinal Diseño evolutivo en el que se estudia la edad de forma intrasujeto, de tal modo que se mide la variable dependiente al mismo grupo en todos los valores de edad.

diseño evolutivo secuencial Diseño evolutivo en el que además de la edad, estudiada longitudinalmente, se estudia el efecto de cohorte.

diseño evolutivo transversal Diseño evolutivo en el que la edad es estudiada de forma intersujeto, de tal modo que la variable dependiente se mide en un grupo diferente de cada edad.

diseño *ex post facto* Diseño que se realiza en unas circunstancias en las que, bien la variable independiente, bien la VI y la VD, ya han tomado sus valores antes de comenzar la investigación.

diseño *ex post facto* prospectivo Diseño *ex post facto* en el que la VI ha tomado ya sus valores. Eso hace que el investigador no pueda manipularla. Se selecciona a los participantes por sus valores en las variables independientes y después se los mide en la variable dependiente.

diseño *ex post facto* retrospectivo Diseño *ex post facto* en el que todas las variables –dependientes e independientes– ya han tomado su valor cuando comienza la investigación. Ésta sólo puede tratar de reconstruir los hechos. Para ello se mide primero la variable dependiente y después, hacia atrás, se buscan posibles variables independientes.

diseño factorial Tipo de metodología experimental con más de una variable independiente. Se caracteriza porque se crean tantas condiciones experimentales como genera la combinación de todos los niveles de todas las variables independientes.

diseño intrasujeto completo Tipo de diseño (experimental) en el que todos los participantes, además de pasar por todas las condiciones experimentales, lo hacen en todas las ordenaciones planeadas en el diseño.

diseño intrasujeto incompleto Tipo de diseño (experimental) en el que los participantes, aun pasando por todas las condiciones experimentales, son divididos en subgrupos, pasando cada subgrupo por una ordenación de las planeadas en el diseño.

diseño longitudinal Plan de investigación, en el contexto de la metodología con encuestas, que consiste en estudiar los cambios producidos en una población a lo largo de un tiempo determinado. Las unidades muestrales son distintas en cada momento del estudio.

diseño transversal Plan de investigación, en el contexto de la metodología con encuestas, que consiste en estudiar las características de una población en un momento dado.

efecto de cohorte Efecto que se produce en una variable dependiente debido a un hecho temporal asociado a los sujetos, como por ejemplo haber nacido en determinada fecha, comenzar unos estudios, ser admitido en una empresa, etc. Se puede enmascarar con el efecto de la edad.

efecto de la práctica En el contexto de los experimentos intrasujeto, ligero cambio en la conducta debido al hecho de haberlos realizado antes en otras condiciones. Para evitar que se acumule de forma sesgada se controla mediante un adecuado control del orden de las condiciones experimentales.

efecto debido al orden véase efecto de la práctica.

efecto principal En el contexto de los diseños factoriales o, en general, en los diseños con más de una variable independiente, acción de una de las variables independientes, sin tener en cuenta su combinación con el resto de las variables independientes. La forma de evaluarlo es mediante los promedios de la variable dependiente en los niveles de la independiente.

efecto simple En el contexto de los diseños factoriales o, en general, en los diseños con más de una variable independiente, cada una de las acciones de una de las variables independientes sobre la dependiente, cuando se combina con cada uno de los niveles de otra de las variables independientes. La forma de evaluarlo es mediante los promedios de la variable dependiente en los niveles de la independiente, combinada con un nivel de la otra variable.

eliminación Técnica de control de las variables extrañas asociadas a las diferencias individuales que consiste en dejar fuera del experimento a los individuos que tengan determinados valores en la variable. Estos valores deben ser muy extremos o distorsionantes respecto a la generalidad.

epistemología Disciplina filosófica relativa al estudio del proceso mediante el cual los hombres generan su conocimiento (episteme).

equilibrado Técnica de control de las variables extrañas asociadas a las diferencias individuales que consiste en hacer que los valores de la variable extraña queden repartidos entre los grupos del experimento. La forma de conseguirlo es mediante la asignación aleatoria de los participantes.

error muestral Margen de error con el que se realizan las estimaciones de la población a partir de los datos descritos en la muestra.

estimación Acción de asignar a la población los valores encontrados en la muestra.

estrato Subconjunto de la población determinado por una variable, cuya relevancia para un objetivo concreto exige que sean reflejados en la muestra.

estudio de panel Plan de investigación, en el contexto de la metodología con encuestas, que consiste en estudiar los cambios producidos en una población a lo largo de un tiempo determinado; las unidades muestrales son las mismas en cada momento del estudio.

experimento Metodología de investigación que se caracteriza por poner a prueba relaciones causales. En un experimento, el investigador reproduce en un contexto controlado una determinada acción de una variable independiente sobre una dependiente. El investigador debe tener la capacidad de generar las condiciones que produzcan los distintos niveles de la variable independiente, así como poder asignar libremente los participantes a las distintas condiciones. Plan de investigación que se caracteriza porque persigue la contrastación de una hipótesis de relación causal entre dos variables y se lleva a cabo en situaciones muy controladas.

experimento con distintos participantes véase diseño de grupos aleatorios.

experimento con los mismos participantes Tipo de experimento en el que todos los participantes pasan por todas las condiciones experimentales. También se les denomina *intrasujetos* o *de medidas repetidas*.

experimento de campo Investigación que contrasta una relación causal entre variables en el contexto natural en que se produce y utilizando la observación como medida de la variable dependiente.

falsación Proceso lógico de razonamiento por medio del cual se buscan datos que vayan en contra de la hipótesis que se postula.

fase de intervención (o tratamiento) Periodo de tiempo, en el contexto de los diseños experimentales con un solo participante, en el que el sujeto recibe algún tipo de intervención por parte del investigador.

fase de retirada En el contexto de los diseños experimentales con un solo sujeto, periodo de tiempo en el que el sujeto no recibe ningún tratamiento, tras haber recibido la intervención por parte del investigador.

fatiga Respuesta del organismo, tras una demanda larga o intensa, caracterizada por un cambio en la conducta que se está midiendo. Antes de empezar, hay que asegurarse de que un efecto alto de la fatiga no invalide el experimento intrasujeto.

fiabilidad Criterio de evaluación de un sistema de recogida de información que nos indica su grado de confiabilidad. Éste puede establecerse mediante acuerdo entre investigadores o mediante el análisis de la estabilidad temporal de la información que nos reporta. Aplicada al conjunto de una investigación, propiedad caracterizada por la constancia de los resultados obtenidos.

frecuencia Tipo de medida que nos informa del número de veces que ha aparecido una categoría dentro de un intervalo de observación.

grupo control Grupo que recibe un nivel nulo de la variable independiente y cuya medida sirve para evaluar los valores del resto de los grupos experimentales.

grupo de cuasi control Grupo de control que es imposible formar mediante la asignación aleatoria de los participantes. También se le conoce como *grupo de control no equivalente*.

grupo experimental Grupo que recibe un nivel no nulo de la variable independiente.

grupos equivalentes Grupos que se han formado mediante la asignación aleatoria de los participantes. Los grupos equivalentes deben tener valores de promedio iguales en la variable dependiente antes de empezar el experimento.

hipótesis Proposición tentativa que se postula, bien como solución a un problema, bien como potencial ley para la explicación de las regularidades observadas.

hipótesis alternativa Proposición que recoge los resultados esperados por el investigador y que es lógicamente complementaria de la hipótesis nula.

hipótesis nula Proposición que recoge los resultados que negarían lo esperado por el investigador, y en la cual el estadístico con que se mide tiene una distribución de probabilidad conocida.

hipótesis operativa Proposición que se expresa en unos términos tales que puede ser contrastada empíricamente y replicada por cualquier otro investigador.

historia Amenaza a la validez interna que puede aparecer en los diseños que necesitan un largo periodo para su implantación. Implica que algunos acontecimientos externos a la investigación pueden producir el mismo efecto que el tratamiento cuyo impacto se investiga y, por tanto, convertirse en una variable enmascarada.

igualación Técnica de control de las variables extrañas asociadas a las diferencias individuales que consiste en hacer que los valores de la variable extraña tengan el mismo valor en todos los participantes.

inconmesurabilidad Propiedad que T.S. Khun postuló para los paradigmas rivales y que implica que sus características no pueden ser reducidas a una métrica común mediante la cual dirimir su rivalidad.

informe de investigación Documento público en el que el investigador detalla todos los pasos del proceso de investigación, presenta los datos obtenidos y sus implicaciones para el objetivo de la investigación.

instrumentación Amenaza a la validez interna que puede aparecer en los diseños que necesitan medir repetidamente las mismas variables. Consiste en que los cambios producidos en las propias pruebas –deterioro, formas no exactamente equivalentes, etc.– se enmascaren con el efecto del tratamiento que se investiga.

intersujeto véase diseño de grupos aleatorios.

interacción Cambio en la forma en que una variable independiente actúa sobre la dependiente en función del nivel de una segunda variable independiente con el que se combina. Cuando existe interacción, los efectos simples de la variable independiente siguen el mismo patrón.

intrasujeto véase *experimento con los mismos participantes*.

ley científica Formulación, verbal o matemática, de las regularidades observadas en los fenómenos que se estudian.

línea base Registro, en el contexto de los diseños experimentales con un solo sujeto, del periodo de tiempo anterior a la intervención por parte del investigador.

maduración Amenaza a la validez interna que puede aparecer en los diseños que necesitan un largo periodo para su implantación. Consiste en que los cambios naturales del organismo que se investiga produzcan un efecto que se enmascare con el efecto del tratamiento.

marco teórico Teoría, o grupo de teorías, desde la que se interpreta el significado de las categorías de la observación.

método deductivo Procedimiento de investigación que genera datos de arriba abajo. Partiendo de una ley, se deriva una implicación contrastable de forma empírica.

método hipotético-deductivo Combinación de las estrategias inductivas y deductivas de investigación en un único proceso. Es el método que usa la Psicología.

método inductivo Procedimiento de investigación que genera conocimiento de abajo arriba. A partir de la observación se detectan regularidades que –una vez contrastadas repetidamente– pasan a postularse como leyes científicas.

metodología cualitativa En su sentido más general, estrategia de investigación de corte subjetivista que persigue la comprensión de lo particular de los fenómenos humanos. Por ello utiliza diseños y técnicas de recogida y análisis de datos de tipo abierto y no estructurado.

metodología cuantitativa En su sentido más general, estrategia de investigación de corte objetivista que persigue el establecimiento de las leyes generales de la naturaleza y los fenómenos humanos. Para ello utiliza diseños y técnicas de recogida y análisis de datos estructurados y de carácter numérico.

metodología de encuestas Tipo de investigación en el que se describen las características de una población, mediante las respuestas de una muestra a un cuestionario elaborado por el investigador.

metodología experimental véase experimento.

motivación En el contexto de la investigación, activación adecuada del comportamiento para realizar las tareas demandadas. Hay que asegurarse la correcta motivación antes de empezar en los diseños *intra*, dado que todos los sujetos deben pasar por todas las condiciones, lo cual podría ser demasiado largo o tedioso.

muestra En el contexto de la metodología de encuestas, conjunto perteneciente a una población a partir de cuyo estudio deseamos extrapolar las características de la población.

muestra representativa Muestra que contiene las características de la población que son relevantes para los objetivos de la investigación.

muestreo Acción de extraer un conjunto de unidades muestrales –muestra– de una población, para realizar una investigación.

muestreo accidental Acción de formar una muestra mediante un accidente que liga las unidades muestrales; la muestra así formada se considera no representativa de la población.

muestreo aleatorio simple Tipo de muestreo probabilístico en el cual las unidades muestrales se seleccionan de forma individual, atendiendo a la probabilidad que las liga a la población.

muestreo estratificado Tipo de muestreo probabilístico en el cual se asegura que los estratos de una o más variables son representados en la muestra.

muestreo a propósito Tipo de muestreo no probabilístico donde las unidades muestrales se incluyen según el criterio de un experto.

muestreo por conglomerados y etapas Tipo de muestreo probabilístico en el que, a lo largo de varias fases –etapas–, se llevan a cabo selecciones de uno o más conglomerados.

nivel véase condición.

nivel de análisis Grado de molaridad y molecularidad en el que se focaliza un sistema de categorías de observación.

nivel de confianza véase confianza.

observación Diseño con objetivo descriptivo que se fundamenta en la codificación de indicadores accesibles a un investigador ajeno al fenómeno que se estudia. También se utiliza este término para referirse a las técnicas de recogida de información basadas en esta estrategia.

observación científica Observación que, además de tener el propósito de aumentar el conocimiento dentro de alguna disciplina, se lleva a cabo de manera sistemática y replicable.

observación estructurada Observación que se lleva a cabo en el lugar natural donde se produce el fenómeno y en la cual el investigador introduce algún tipo de restricción en la situación con el fin de aumentar sus posibilidades de aparición.

observación externa Observación en la que el investigador no forma parte del fenómeno que se observa.

observación natural Observación que se lleva a cabo en el lugar natural donde se produce el fenómeno y en la que el observador no interviene de ninguna manera.

observación participante Observación que se lleva a cabo desde dentro mismo del fenómeno: el observador es parte de lo que se observa.

observador ciego Observador que no conoce el propósito de la investigación para la que recoge los datos.

ocurrencia Tipo de medida de la observación que informa de la aparición de una determinada categoría en un determinado intervalo de observación.

operativización de variables Acción de definir las variables de forma que cualquier investigador las pueda observar y medir en cualquier momento.

ordenación En el contexto de los experimentos intrasujeto, secuencia completa de todas las condiciones. La posición de las condiciones dentro de la ordenación se hace atendiendo al tipo de control del orden que se realiza. Se forman tantas ordenaciones como repeticiones se desean.

ordenación aleatoria en bloques Procedimiento de control del efecto de la práctica en los diseños intrasujeto. Cada bloque es una ordenación aleatoria del total de las condiciones. Se preparan tantos bloques como repeticiones son necesarias.

ordenación aleatoria simple Procedimiento de control del efecto de la práctica en los diseños intrasujeto. Se caracteriza por presentar de forma aleatoria el total de las presentaciones de las condiciones experimentales.

paradigma Conjunto de teorías dentro de una disciplina, organizadas de modo tal que implican una definición de su ámbito y de su metodología. También implica un conjunto de reglas implícitas para la relación entre los investigadores que se adscriben a él.

pérdida no aleatoria de sujetos Amenaza a la validez interna consistente en que los sujetos que abandonan la investigación lo hagan por razones que tienen que ver con la propia investigación, de tal modo que los resultados obtenidos no serían los mismos si hubiesen permanecido.

persistencia En el contexto de los experimentos intrasujeto, resto del efecto de una condición cuando se presenta la siguiente. Se controla dejando pasar el tiempo adecuado o con una adecuada ordenación de las condiciones.

persistencia diferencial En el contexto de los experimentos intrasujeto, efecto de un nivel sobre el consecutivo de tal naturaleza que no es posible reequilibrarlo con la ordenación inversa. Cuando ocurre, el experimento se invalida.

población En el contexto de las metodologías con encuestas, conjunto que se desea describir en alguna o algunas de sus variables, pero cuyo tamaño o dispersión hacen muy difícil la tarea.

precisión En las estimaciones de los valores de la población, amplitud determinada por la suma y resta del error muestral, relativa al rango de la variable estudiada.

pregunta de respuesta abierta En el contexto de las metodologías con encuestas, pregunta del cuestionario, a la cual puede responder el encuestado lo que le parezca.

pregunta de respuesta cerrada En el contexto de las metodologías con encuestas, pregunta del cuestionario a la cual el encuestado sólo puede responder eligiendo una de las respuestas presentadas.

problema de investigación Zona oscura del conocimiento científico establecido caracterizada, bien por la necesidad de explorar un nuevo ámbito de investigación, bien por la existencia de datos contradictorios en cuanto al funcionamiento de una ley ya establecida.

procedimiento Apartado del informe de investigación en el que se detallan cuidadosamente todos y cada uno de los pasos del diseño, de investigación desarrollado en un estudio particular.

proceso de investigación Secuencia de pasos que empieza por el planteamiento de un problema, la formulación de una hipótesis, el diseño de un plan de investigación, la recogida de datos una vez aplicado el mismo y la conclusión acerca de las implicaciones de éstos para el problema con que se inició la mencionada secuencia.

reactividad Cambio en la conducta del sujeto observado por el simple hecho de saber que se le observa.

registro narrativo Tipo de registro de la observación en el que se usa el lenguaje como modo de dejar constancia de lo observado.

registro por intervalos Procedimiento de registro consistente en fragmentar los intervalos de observación en unidades muy breves intercalando el momento de la observación con el momento del registro.

regresión a la media Tendencia de algunas puntuaciones extremas consistente en perder su carácter extremo cuando se repite la medida. Se puede enmascarar con el efecto del tratamiento en las investigaciones que seleccionan a los participantes por poseer un valor extremo en la variable independiente y cuya hipótesis es que el tratamiento hará que las puntuaciones no sean extremas.

relación espuria Relación observada entre dos variables pero que no implica vínculo entre ellas, ya que se produce al estar ambas relacionadas con una tercera variable.

repetición En el contexto de los experimentos intrasujeto, operación que se da cada vez que se presenta la misma condición experimental para ser evaluada en los participantes. Las repeticiones son necesarias para poder registrar el efecto de la variable independiente.

replicabilidad Característica del modo de generar el conocimiento de la ciencia que permite que cada nueva aportación pueda ser repetidamente contrastada por cualquier investigador.

ruta aleatoria Procedimiento, en el contexto de las metodologías con encuestas, para seleccionar la unidad muestral a partir del último conglomerado. Consiste en ir sorteando cada uno de los valores de las variables que definen su ubicación: calle, bloque, número, planta y letra.

sensibilidad En el contexto de la experimentación, cualidad de la operativización de las variables, de la tarea y de los aparatos de medida, para registrar la presencia del fenómeno en estudio.

sesgo Mala representación de la población en la muestra estudiada en el contexto de las metodologías con encuestas.

sesgo de respuesta Influencia en las respuestas de los encuestados derivada del hecho de poder conocer de antemano las preguntas antes de empezar a contestar.

sesgo de selección Amenaza a la validez interna producida por la formación de grupos de modo no aleatorio. Cuando se selecciona a un sujeto por una característica, no se la puede aislar de otras que también puedan estar influyendo en la variable dependiente y que, de ser así, quedarían enmascaradas.

sesgo del entrevistador Influencia parcial del entrevistador sobre el entrevistado, o sobre la codificación de las respuestas del entrevistado.

sesgo retrospectivo Amenaza a la validez interna que puede aparecer en los diseños *ex post facto* retrospectivos por el hecho de que el investigador puede dejarse variables relevantes sin medir en su búsqueda hacia atrás de potenciales variables independientes.

tamaño de la muestra En el contexto de las metodologías con encuesta, número de unidades muestrales necesarias para que una muestra sea representativa respecto a una determinada población.

tarea Conjunto de acciones que deben ser llevadas a cabo por los participantes en cada condición experimental. Son planeadas por el investigador según la operativización de la variable independiente y dependiente. Por medio de ellas se espera que se ponga de manifiesto el fenómeno en estudio.

técnica de control del orden Técnica empleada en el contexto de los experimentos intrasujeto para contrarrestar los ineludibles efectos de la práctica. Son técnicas de control del orden: la ordenación aleatoria simple, la ordenación aleatoria en bloques, el reequilibrado y el cuadrado latino.

teoría Conjunto de leyes interrelacionadas que hacen referencia al mismo ámbito disciplinario.

unidad muestral Elemento de la población que genera la información objetivo del estudio. Puede ser un individuo físico o una agrupación de individuos definida por cualesquiera criterios.

validez Criterio de evaluación del proceso de investigación que nos informa del grado en que se ha conseguido realmente el propósito que se buscaba. En el caso de la observación, la validez indica el grado en que se ha observado lo que se pretendía y no otra cosa.

validez de constructo Cuando un conjunto de categorías se postula como indicadores de un constructo no accesible a la observación directa, la validez de constructo establece el grado en que estos indicadores cumplen su propósito.

validez de contenido Variante del concepto de validez que hace referencia al grado en que el conjunto de categorías incluidas en el código constituye una muestra representativa de todos sus elementos.

validez externa Cualidad de la investigación que permite generalizar los resultados encontrados al resto de los individuos y las situaciones que no fueron estudiados.

variable controlada Antigua variable extraña cuyo efecto sobre la dependiente ha sido controlado por el investigador.

variable dependiente Variable cuyos cambios tratan de ser explicados por las variaciones en la variable independiente. Cuando se trata de metodologías no experimentales, se explican por relaciones o diferencias entre grupos; cuando se trata de metodologías experimentales, mediante relaciones causales.

variable enmascarada Tipo de variable extraña cuya acción sobre la dependiente tiene la característica de que varía al mismo tiempo que varían los niveles de la independiente y, además, el efecto es tal que podría explicar las variaciones en la dependiente.

variable extraña Variable que afecta a la dependiente y que no ha sido ni manipulada (independiente) ni controlada. En una investigación bien controlada, la cuantía del efecto de estas variables es siempre menor que el efecto de las manipuladas.

variable independiente Variable cuyos cambios tratan de explicar las variaciones en la variable dependiente. Cuando se trata de metodologías no experimentales, el investigador sólo puede medir y seleccionar valores de la variable. Cuando se trata de metodologías experimentales, el investigador puede crear las condiciones para que se manifieste y aplicarla a los participantes según su deseo (manipularla).

variable manipulada véase variable independiente.

verificación Estrategia lógica de contraste de hipótesis consistente en la búsqueda de datos que la apoyen.

