



Experimentos

Lo que aprenderá en este capítulo

En este capítulo examinaremos el método experimental. Además le aprender un modo particular de observación, apreciará una lógica de la indagación que puede aplicar también a otros modos.

En este capítulo...

Introducción

Temas apropiados para efectuar experimentos

El experimento clásico

Variabes independiente y dependiente

Pretest y postest

Grupo experimental y control

El experimento doble ciego

Selección de los sujetos

Muestreo probabilístico

Aleatorización

Igualamiento

Variaciones del diseño experimental

Fuentes de invalidez interna

Fuentes de invalidez externa

Ejemplo de experimentos

Grupos focales o de interés

Experimentos "naturales"

Ventajas y desventajas del método experimental

Puntos principales

Preguntas y ejercicios de repaso

Proyecto de continuidad

Lecturas adicionales

Introducción

Este capítulo aborda el método de investigación que más se asocia con la ciencia estructurada en general. Aquí estudiaremos la *experimentación* como forma de observación científica. Básicamente, los experimentos consisten en (1) emprender una acción y (2) observar las consecuencias de tal acción. Los científicos sociales acostumbran elegir un grupo de sujetos, hacer algo con ellos y observar el efecto. En este capítulo, examinaremos tanto la lógica como las técnicas de los experimentos científicos sociales.

Conviene anotar desde el principio que los experimentos también se realizan en la indagación no científica. Por ejemplo, al preparar un estofado ponemos sal, probamos, añadimos más sal y probamos de nuevo. Al desactivar una bomba, cortamos un cable, observamos si explota, cortamos otro cable, y...

También experimentamos abundantemente en nuestros esfuerzos por adquirir una comprensión general del mundo en que vivimos. Todas nuestras habilidades se aprenden por experimentación: comer, caminar, hablar, andar en bicicleta, nadar, etc. Por experimentación los alumnos descubren cuánto tienen que estudiar para obtener éxito aca-

démico. Por experimentación los profesores aprenden cuánta preparación se necesita para ofrecer buenas lecciones. En este capítulo analizaremos algunas formas en que los científicos sociales realizan experimentos para obtener conocimientos generalizados. Veremos que, al igual que otros métodos de las ciencias sociales, experimentar tiene sus ventajas y sus inconvenientes.

Temas apropiados para efectuar experimentos

Los experimentos se prestan especialmente bien para los proyectos de investigación que comprenden proposiciones y conceptos relativamente limitados y bien definidos. La imagen tradicional de las ciencias, que recordamos al comienzo del libro, y el *modelo experimental* están muy relacionados. Así, la experimentación es en particular conveniente para someter a prueba hipótesis. También se adapta mejor a los propósitos explicativos que a los descriptivos. Por ejemplo, supongamos que queremos estudiar los prejuicios contra los negros y descubrir los medios para aminorarlos. Planteamos la hipótesis de que el hecho de conocer la aportación de los negros a la historia mundial tendrá el efecto

de reducir los prejuicios, y la ponemos a prueba con un experimento. Para empezar, podríamos examinar a un grupo de sujetos experimentales para determinar el grado de sus prejuicios contra los negros. Acto seguido, les proyectaríamos un documental de las contribuciones importantes de los negros al avance científico, literario, político y social del mundo. Por último, mediríamos de nuevo el grado de prejuicios de nuestros sujetos contra los negros para determinar si el filme en realidad los redujo.

La experimentación también es apropiada y ha tenido éxito para estudiar las relaciones personales en grupos reducidos. Así, podríamos reunir un grupo pequeño de sujetos experimentales y les asignaríamos una tarea, como hacer recomendaciones para compartir el automóvil. Entonces, observaríamos la manera en que se organiza el grupo y enfrenta el problema. En el transcurso de varios experimentos podríamos variar la naturaleza de la tarea o las recompensas por resolverla. Al observar las diferencias de la forma en que los grupos se organizan y funcionan en esas circunstancias variadas, aprendemos mucho sobre el carácter de las interacciones en los grupos pequeños.

Habitualmente pensamos que los experimentos se realizan en laboratorios, y, en efecto, la mayor parte de los ejemplos que citaremos en el capítulo se refieren a ese entorno; sin embargo, no tiene que ser así por fuerza. Como veremos, los científicos sociales suelen estudiar lo que llaman *experimentos naturales*, "experimentos" que ocurren en el curso normal de los sucesos sociales. Dedicaremos la última parte del capítulo a estas investigaciones.

El experimento clásico

La clase más convencional de experimento, tanto en las ciencias naturales como en las sociales, comprende tres pares principales de componentes: (1) variables independiente y dependiente, (2) pretest y postest, y (3) grupos experimental y control. En esta sección del capítulo nos ocuparemos de los tres y de la forma en que se conjuntan para la realización del experimento.

Variables independiente y dependiente

En esencia, un experimento indaga el efecto de una variable independiente en una variable dependien-

te. En general, la variable independiente adopta la forma de un estímulo experimental, que está presente o bien falta; es decir, se trata de una *variable dicotómica*, con dos atributos (pero, como indicaremos en secciones posteriores, no siempre tiene que ser así). En el ejemplo de los prejuicios contra los negros, *prejuicios* es la variable dependiente y *exposición a la historia de los negros* la independiente. Según la hipótesis del investigador, los prejuicios dependen en parte de una falta de conocimientos sobre la historia de los negros. El propósito del experimento es probar la validez de tal hipótesis. Las variables independientes y dependientes que se prestan a la experimentación son casi ilimitadas. Más aún, una variable puede ser la independiente en un estudio y la dependiente en otro. Así, los *prejuicios* son la variable dependiente en el ejemplo anterior, pero podrían ser la variable independiente de un experimento que indagara el efecto de los prejuicios en la emisión de votos.

En otros términos, la variable independiente es la causa y la dependiente el efecto. Por tanto, podríamos decir que observar el filme provocó un cambio en los prejuicios, o que la mengua de los prejuicios fue un efecto de mirar el documental.

Tanto la variable independiente como la dependiente deben definirse en forma operacional para fines de la experimentación. Estas definiciones operacionales pueden comprender varios métodos de observación. Por ejemplo, las respuestas de un cuestionario pueden ser la base para definir los prejuicios. Los elementos de la definición operacional de las relaciones con los negros en grupos pequeños pueden ser: hablar con los sujetos negros, estar de acuerdo con ellos, disentir con ellos o ignorar sus comentarios.

De manera convencional, en el modelo experimental las variables dependiente e independiente deben contar con una definición operacional antes de comenzar el experimento. Sin embargo, como veremos a propósito de las encuestas y otros métodos de investigación, a veces es apropiado hacer primero una amplia variedad de observaciones durante la recopilación de datos y en un análisis posterior determinar las definiciones operacionales más útiles. Como quiera que sea, a fin de cuentas la experimentación, como otros métodos cuantitativos, requiere mediciones y observaciones concretas y estandarizadas.

Pretest y postest

En el diseño experimental más simple, se miden los sujetos según una variable dependiente (pretest), se exponen a un estímulo que representa la variable independiente y se miden de nuevo de acuerdo con la variable dependiente (postest). Las diferencias que se adviertan entre la primera y la última medición de la variable dependiente se atribuyen a la variable independiente.

En el ejemplo de los prejuicios y la exposición a la historia de los negros, comenzamos con un pretest del grado de prejuicios entre los sujetos de nuestro experimento. Por ejemplo, con un cuestionario que los interroga sobre sus actitudes hacia los negros, mediríamos los prejuicios que exhibe cada sujeto y el nivel promedio de prejuicios de todo el grupo. Después de hacerlos presenciar el filme con la historia de los negros, les daríamos el mismo cuestionario. Las respuestas al postest nos permitirían medir el nuevo grado de prejuicios de cada individuo y el promedio de la totalidad del grupo. Si descubrimos un nivel menor de prejuicios durante la segunda aplicación del cuestionario, concluiríamos que en efecto el documental aminó los prejuicios.

En el examen experimental de actitudes de este tipo, como los prejuicios, enfrentamos un problema práctico especial que se relaciona con la *validez*. Como usted ya se habrá imaginado, es posible que los sujetos respondan de manera distinta a los cuestionarios la segunda vez aunque no hayan cambiado sus actitudes. Durante la primera aplicación del cuestionario, quizá los sujetos no eran conscientes del objetivo. Para el momento de la segunda medición, habrían inferido que a los investigadores les interesa medir sus prejuicios. Como nadie quiere parecer prejuicioso, los sujetos podrían "limpiar" sus respuestas la segunda vez. Así, parecería que el documental redujo los prejuicios cuando en realidad no lo hizo.

Este es un ejemplo de un problema general que abunda en muchas formas de investigación social científica. El solo hecho de estudiar algo puede alterarlo. En varias partes del capítulo revisaremos las técnicas para enfrentar este problema en el contexto de la experimentación.

Grupos experimental y control

El primer método para compensar los efectos del propio experimento es el establecimiento de un *grupo control*. Los experimentos de laboratorio rara vez o nunca se reducen a la observación del grupo experimental al que se presenta el estímulo; además, los investigadores también observan a un grupo control que no recibe el estímulo experimental.

En el ejemplo de los prejuicios y la historia de los negros, se examinarían dos grupos de sujetos. Para empezar, cada grupo contesta un cuestionario destinado a medir sus prejuicios contra los negros. En seguida, el grupo experimental ve el filme. Después, el investigador aplica un postest de prejuicios a ambos grupos. La figura 9.1 ilustra el diseño experimental básico.

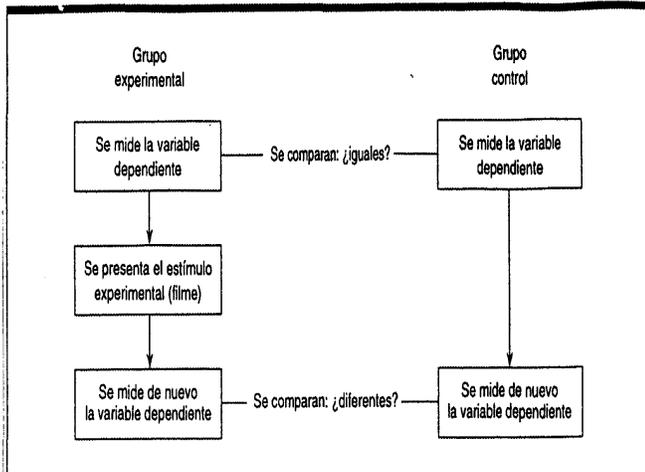
El hecho de tener un grupo control permite al investigador detectar cualesquiera efectos del propio experimento. Si el postest muestra que el nivel general de prejuicios del grupo control se ha reducido tanto como en el grupo experimental, esta reducción manifiesta debe ser producto del experimento o de algún factor externo, y no un resultado del documental. Por otra parte, si los prejuicios disminuyeron sólo en el grupo experimental, la reducción debe ser consecuencia de la exhibición del filme, puesto que es la única diferencia entre los dos grupos. Asimismo, si los prejuicios se redujeron más en el grupo experimental que en el control, también habría fundamentos para suponer que el filme es la causa.

La necesidad de grupos control en la investigación social se aclara en una serie de estudios de la satisfacción laboral que realizaron F. J. Roethlisberger y W. J. Dickson (1939) a finales de la década de 1920 y principios de la de 1930. Estos dos investigadores estudiaron las condiciones de trabajo en el "almacén de cables" telefónicos de la Western Electric Works en Hawthorne, en los suburbios de Chicago, Illinois, para tratar de descubrir los cambios de condiciones laborales que mejorarían la satisfacción de los empleados y la productividad.

Para gran satisfacción de los investigadores, descubrieron que al mejorar las condiciones laborales se incrementaba la satisfacción y la productividad. Por ejemplo, cuando aumentaron la iluminación del lugar de trabajo, la productividad creció; al au-

Figura 9.1

Diagrama del diseño experimental básico



mentaría de nuevo, la productividad creció otra vez. Para justificar su conclusión científica, los investigadores disminuyeron las luces: la productividad ascendió una vez más.

Se hizo evidente que los trabajadores de la sala de cables respondían más a la atención que les brindaban los investigadores que a la mejora de las condiciones laborales. Como resultado de este fenómeno, denominado **efecto Hawthorne**, los investigadores procuran ser más sensibles y precavidos ante los efectos posibles de los propios experimentos. La comparación con un grupo control adecuado examinado a fondo sin que por lo demás cambiaran las condiciones de trabajo habría indicado la presencia de este efecto en el estudio de la sala de cables.

En ningún campo es más evidente la necesidad de grupos control en la experimentación que en la investigación médica. Una y otra vez, parece que los pacientes que participan en los experimentos médicos mejoran sin que quede claro cuánto de la mejora se debe al tratamiento experimental y cuánto al experimento. Por eso, al probar los efectos de los medicamentos nuevos, los investigadores médicos acostumbra dar un *placebo* (por ejemplo, pastillas de azúcar) a un grupo control. Así, los pacientes del grupo control creen que, al igual que el grupo experimental, toman un fármaco en estudio.

Con frecuencia mejoran. Sin embargo, si el medicamento nuevo es eficaz, quienes lo toman mejoran más que los que ingieren placebos.

En los experimentos sociales científicos, los grupos control son importantes para precavernos no sólo de los efectos de los experimentos, sino también de sucesos externos al laboratorio que ocurren durante los experimentos. En el ejemplo del estudio de los prejuicios, supongamos que un líder negro muy popular fuera asesinado a la mitad de un experimento de, digamos, una semana de duración. El acontecimiento horrorizaría a los sujetos y los obligaría a revisar sus actitudes hacia los negros, con el resultado de que se reducirían sus prejuicios. Como este efecto debería incidir por igual en los miembros de los grupos experimental y control, una reducción de los prejuicios mayor entre los primeros señalaría el impacto del estímulo: la película documental.

A veces, un diseño experimental requiere más de un grupo experimental o control. Por ejemplo, en el ejemplo de la película documental, tal vez queramos examinar también el efecto de la lectura de un libro sobre la historia de los negros. En tal caso, haríamos que un grupo viera el filme y leyera el libro, otro sólo vería la película, uno más leería el libro y el grupo control no haría lo uno ni lo otro. Con este diseño determinaríamos el impacto de ca-

da estímulo por separado, así como su efecto combinado.

El experimento doble ciego

Al igual que los pacientes que mejoran cuando creen que toman un medicamento nuevo, a veces los experimentadores tienden a prejujgar los resultados. En la investigación médica, es más probable que los investigadores "observen" mejoras en los pacientes que reciben el fármaco experimental que en quienes toman el placebo (y suele ser más probable si los mismos investigadores sintetizaron el medicamento). Un experimento doble ciego elimina esta posibilidad, porque ni los sujetos ni los investigadores saben cuál es el grupo experimental ni cuál el control. En el caso médico, no se informaría a los investigadores encargados de administrar el fármaco y vigilar las mejoras que sujetos lo toman y cuáles toman el placebo. Al mismo tiempo, el investigador que supiera en qué grupo está cada paciente no ejecutaría el experimento.

En los experimentos sociales científicos, al igual que en los médicos, el peligro del sesgo del experimentador se reduce más en la medida en que las definiciones operacionales de las variables dependientes son claras y precisas. Así, sería menos probable que los investigadores médicos sesgaran inconscientemente su lectura de la temperatura de un paciente que su evaluación del grado de letargo del enfermo. Por la misma razón, el investigador de un grupo pequeño tiene menos probabilidad de tergiversar su percepción al detectar qué sujeto habla o a quién se dirige que al dilucidar si sus comentarios suenan participativos o competitivos.

Como hemos dicho varias veces, casi nunca podemos establecer definiciones y mediciones operacionales completamente precisas e inequívocas. Por tanto, en ocasiones conviene emplear un diseño doble ciego en los experimentos de investigación social.

Selección de los sujetos

Es de creer que la mayoría de los experimentos sociales de laboratorio se realizan con estudiantes de licenciatura como sujetos. Habitualmente, el investigador les pide a los alumnos inscritos en su clase

que participen en experimentos o bien los solicita en el periódico escolar. Los sujetos pueden o no recibir una paga por su colaboración en el experimento (en el capítulo 18 estudiaremos las cuestiones éticas que surgen al pedir a los estudiantes que participen en tales estudios).

En relación con la norma científica de la *generalizabilidad*, esta tendencia representa sin duda un posible defecto en la investigación social científica. Para decirlo en forma sencilla, los estudiantes de licenciatura no son característicos del conjunto de la población. Por tanto, se corre el peligro de que aprendamos mucho de las actitudes y los actos de los universitarios, pero nada sobre actitudes y actos en general.

Sin embargo, este posible defecto es menos importante en la investigación explicativa que en la descriptiva. Es verdad que al observar el grado de prejuicios en un grupo de estudiantes de licenciatura estaríamos poco confiados en encontrar el mismo grado en la generalidad de la población. Por otro lado, si resultara que una película documental reduce los prejuicios entre los estudiantes, tendríamos más confianza sin estar seguros en que produciría un efecto parecido en la comunidad. Los procesos sociales y los patrones de relaciones causales son más estables y generalizables que las características específicas.

Aparte del asunto de la generalizabilidad, la regla cardinal de la selección de sujetos y la experimentación atañe a la capacidad de comparar a los grupos experimental y control. En términos ideales, el grupo control representa lo que sería el grupo experimental de no haber estado expuesto al estímulo. Por tanto, es esencial que el grupo experimental y el control sean tan parecidos como sea posible. Hay varios modos de lograrlo.

Muestreo probabilístico

El análisis anterior de la lógica y las técnicas del muestreo probabilístico nos brinda un método para elegir dos grupos similares de personas. Al comenzar con un marco de muestreo compuesto por todas las personas de la población en estudio, el investigador puede elegir dos muestras probabilísticas. Si ambas se asemejan a la población de la que fueron extraídas, también se parecerán una a la otra.

Fuentes de invalidez interna

El problema de la **invalidez interna** se refiere a la posibilidad de que las conclusiones derivadas de los resultados del experimento no reflejen con exactitud lo que sucedió en el propio experimento. La amenaza de invalidez interna está presente siempre que algún factor aparte del estímulo experimental pueda influir en la variable dependiente. Campbell y Stanley (1963:5-6) y T. Cook y Campbell (1979:51-55) señalan varias fuentes del problema. Veamos 12:

1. **Historia.** Pueden ocurrir acontecimientos históricos en el transcurso del experimento que confundan los resultados. Un ejemplo sería el asesinato de un líder negro durante un experimento para reducir los prejuicios contra los negros.

2. **Maduración.** Crecemos y cambiamos de continuo, estemos o no en un experimento, y tales cambios pueden influir en los resultados. En un experimento a largo plazo, el hecho de que los sujetos envejezcan (¿y sean más sabios?) puede tener un efecto. En los experimentos más breves, los sujetos se cansan, duermen, se aburren, les da hambre o cambian de otras maneras que pueden afectar su conducta durante el experimento.

3. **Administración de pruebas.** A menudo la aplicación de pruebas iniciales y pruebas posteriores influirá en el comportamiento de la gente y confundirá los resultados del experimento. Supongamos que damos un cuestionario a un grupo con el fin de medir los prejuicios. En seguida, presentamos un estímulo experimental y medimos de nuevo los prejuicios. Cuando realicemos el posttest, es probable que los sujetos estén más al tanto del tema de los prejuicios y meditarán mejor sus respuestas. De hecho, tal vez infirieron que queremos averiguar cuán prejuiciosos son, y como a casi nadie le gusta parecer prejuicioso, mostrarán su mejor conducta y darán las respuestas que creen que queremos o que los harán verse bien.

4. **Instrumentación.** Hasta aquí, no hemos dicho gran cosa del proceso de medición en los pretest y posttest, pero quiero recordarle los problemas de concepción y operacionalización que ya estudiamos. Si usamos otras medidas de la variable de-

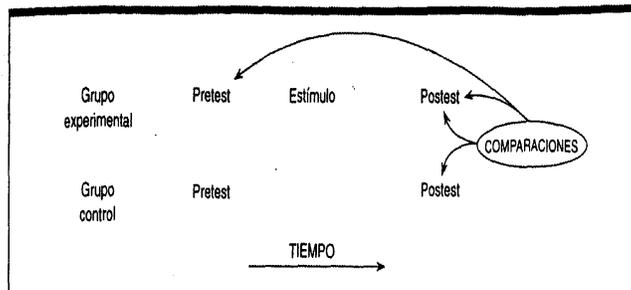
pendiente (digamos, otros cuestionarios sobre los prejuicios), ¿cómo estamos seguros de que son equivalentes? Quizá parezca que disminuyen los prejuicios simplemente porque la medición en el pretest fue más exacta que la medición en el posttest. O, si los experimentadores hacen las mediciones, sus normas o sus capacidades pueden cambiar en el transcurso del experimento.

5. **Regresión estadística.** A veces es apropiado realizar experimentos con sujetos que al principio tienen calificaciones extremas en la variable dependiente. Si quiere probar un nuevo método para enseñar matemáticas a fracasados empedernidos, preferirá realizar su experimento con sujetos que se hayan desempeñado muy mal en la materia. Pero medite un momento en lo que es probable que ocurra al cabo del tiempo con estas personas sin ninguna interferencia experimental. Empezaron tan abajo que sólo pueden quedarse en el fondo o mejorar: no pueden empeorar. Así, es probable que el grupo muestre algunas mejoras con el tiempo, aunque no tenga ningún estímulo experimental. Al hablar de *regresión a la media*, los estadísticos señalan que es probable que las personas extremadamente altas, como grupo, tengan hijos más bajos y que las personas extremadamente bajas también como grupo tengan hijos más altos. Por tanto, se corre el peligro de que los cambios que acontecen en virtud del punto de partida extremo de los sujetos se atribuyan por error a los efectos del estímulo experimental.

6. **Sesgos en la selección.** Ya hablamos de los sesgos en la selección cuando examinamos las formas de elegir a los sujetos para los experimentos y de asignarlos a los grupos experimental y control. Las comparaciones no tendrán ningún sentido a menos que los grupos sean *comparables*.

7. **Mortandad experimental.** Aunque me imagino que algunos experimentos podrían matar a sus sujetos, este problema atañe a una forma más general y menos extrema de mortandad. A menudo, los sujetos abandonan el experimento antes de que termine, lo que puede afectar las comparaciones estadísticas y las conclusiones. En el experimento clásico que comprende un grupo experimental y uno control, ambos con pretest y posttest, supon-

Figura 9.4
Otro esquema del experimento clásico



gamos que los fanáticos del grupo experimental se ofenden tanto con la película de la historia de los negros que le dicen al experimentador que lo olvide y se marchan. Los sujetos que se quedan para el posttest ya eran menos prejuiciosos desde el principio, por lo que los resultados del grupo manifestarán una "disminución" sustancial en los prejuicios.

8. **Orden causal temporal.** Si bien es rara en la investigación social, puede surgir una ambigüedad sobre la secuencia del estímulo experimental y la variable dependiente. Cuando ocurre, la conclusión de que el estímulo causó la variable dependiente puede ponerse en tela de juicio con la explicación de que en realidad la variable "dependiente" causó cambios en el estímulo (véase el capítulo 3).

9. **Difusión o imitación de tratamientos.** Cuando los sujetos del grupo experimental y los del control pueden comunicarse, es posible que los primeros transmitan algunos elementos del estímulo a los segundos. Algunos miembros del grupo experimental podrían comentar a los del grupo control acerca del filme de la historia de los negros. En ese caso, el grupo control queda influido por el estímulo y ya no es un control real. A veces decimos que el grupo control "se contaminó".

10. **Compensación.** Como veremos en el capítulo 13, en los experimentos en situaciones reales como un programa de educación especial se suele privar a los sujetos del grupo control de algo que se considera valioso. En tales casos, puede haber presiones para ofrecerles alguna forma de compensación.

Por ejemplo, el personal de un hospital puede sentir pena por los pacientes del grupo control y darles más "cuidados amorosos". En tal situación, ya no tenemos un grupo control auténtico.

11. **Rivalidad compensatoria.** En los experimentos en situaciones reales, los sujetos privados del estímulo experimental pueden tratar de compensar su falta esforzándose más. Supongamos que un programa experimental de matemáticas es el estímulo; los miembros del grupo control podrían aplicarse más que antes a las matemáticas para vencer a los sujetos "especiales" del grupo experimental.

12. **Desmoralización.** Por otro lado, los sentimientos de privación de los miembros del grupo control pueden llevarlos a renunciar. En los experimentos educativos, los sujetos de grupos control desmoralizados pueden dejar de estudiar, fallar o irritarse.

Éstas, pues, son algunas fuentes de invalidez interna que citaron Campbell, Stanley y Cook. Al tanto de ellas, los experimentadores han ideado diseños destinados a manejarlas. El experimento clásico, que estudiamos al principio del capítulo, si cuenta con una selección y una asignación apropiadas de los sujetos, se ocupa de todos estos problemas. Veamos de nuevo el diseño de estudio en el esquema de la figura 9.4.

Para continuar con el ejemplo del filme con la historia de los negros y el objetivo de disminuir los prejuicios contra ellos, si nos valemos del diseño experimental que se muestra en la figura 9.4, esperaríamos dos resultados. En el grupo experimental, el nivel de prejuicios que se midió en el posttest de-

be ser menor que el del pretest. Además, cuando comparamos las dos postest, debemos encontrar menos prejuicios en el grupo experimental que en el control.

Este diseño también nos previene contra el problema denominado de historia, pues si ocurriera algo ajeno al experimento que pudiese influir en el grupo experimental, también afectaría al grupo control y aun surgiría una diferencia en los resultados de los dos postest. La misma comparación nos evita los problemas de maduración siempre que los sujetos se hayan asignados al azar en los dos grupos. Las pruebas y la instrumentación no pueden ser problemas, puesto que tanto el grupo experimental como el control están sujetos a los mismos efectos de las pruebas y del experimentador. Si los sujetos se asignaron en forma aleatoria a los dos grupos, la regresión estadística los afectaría a ambos por igual, aunque se estudie a personas con posturas radicales de prejuicios (o cualquier otra variable que se examine). La asignación aleatoria de los sujetos descarta los sesgos en la selección. Es más complicado manejar la mortandad experimental, pero los datos que arroja este diseño de estudio brindan varios medios de enfrentarla. Algún tipo de modificaciones ligeras en el diseño —por ejemplo, dar un placebo (como proyectar una película que no tenga nada que ver con los negros) al grupo control— puede facilitar el manejo del problema.

Los cinco problemas restantes de invalidez interna se evitan mediante la aplicación cuidadosa de un diseño experimental controlado. El diseño experimental que estamos estudiando facilita la especificación clara de las variables independiente y dependiente. Los sujetos de los grupos experimental y control deben mantenerse separados para reducir la posibilidad de difusión o imitación de tratamientos. Los controles de aplicación previenen la necesidad de compensar al grupo control y la rivalidad compensatoria se vigila y se tiene en cuenta a la hora de evaluar los resultados del experimento, lo mismo que el problema de la desmoralización.

Fuentes de invalidez externa

La invalidez interna da cuenta de algunas de las complicaciones que enfrentan los experimentadores. Además, hay problemas que Campbell y Stan-

ley llaman de **invalidez externa**, que se relacionan con la **generalizabilidad** de los hallagos del experimento al mundo "real". Aunque los resultados sean una medición exacta de lo que ocurrió durante el experimento, ¿realmente nos dicen algo de la vida en la espesura de la sociedad?

Campbell y Stanley describen cuatro formas de este problema; nosotros presentaremos una a modo de ilustración. Según los autores, la generalizabilidad de los resultados de un experimento está en riesgo si hay una interacción entre la situación de prueba y el estímulo experimental (1963:18). Veamos un ejemplo de lo que quieren decir.

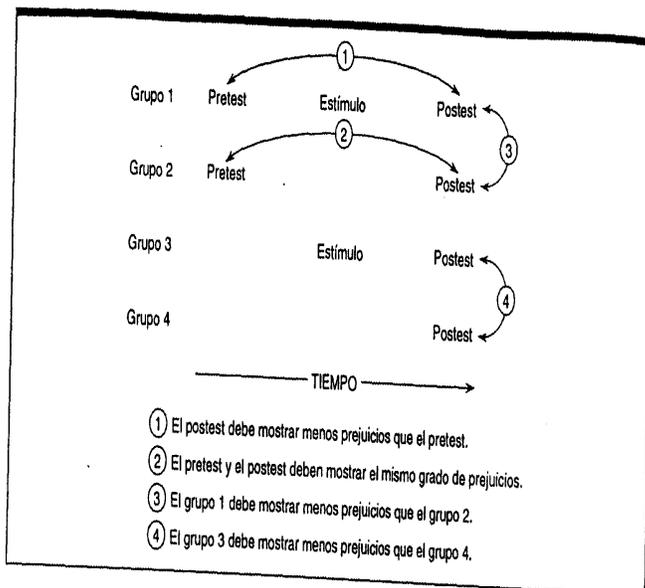
Siempre con el estudio de los prejuicios y la película de la historia de los negros, supongamos que nuestro grupo experimental en el experimento clásico tiene menos prejuicios en el postest que en el pretest, y menos también que el postest del grupo control. Nos sentiríamos confiados en decir que el filme redujo en verdad los prejuicios de los sujetos del grupo experimental. ¿Pero tendría el mismo efecto la película si se proyectara en cines o en la televisión? No podemos estar seguros, porque acaso sólo funcione cuando la gente se haya sensibilizado con el tema de los prejuicios, como pudo haber ocurrido con los sujetos al responder el pretest. Éste es un ejemplo de interacción entre la prueba y el estímulo. El diseño experimental clásico no puede controlar esa posibilidad. Por fortuna, los experimentadores han concebido otros diseños que sí lo logran.

El **diseño de cuatro grupos de Solomon** (Campbell y Stanley, 1963:24-25) aborda el problema de esta interacción de la prueba y el estímulo. Como lo indica su nombre, consta de cuatro grupos de sujetos asignados al azar de un conjunto general. La figura 9.5 presenta un esquema de este diseño.

Observe que los grupos 1 y 2 de la figura 9.5 componen el experimento clásico. El grupo 3 presencia el estímulo experimental sin pretest y el grupo 4 sólo realiza el postest. Este último diseño experimental permite cuatro comparaciones significativas. Si la película con la historia de los negros disminuye realmente los prejuicios, sin que explique la reducción por los problemas de la validez interna y la interacción de prueba y estímulo, debemos esperar cuatro resultados:

1. El grupo 1 debe exhibir menos prejuicios en el postest que en el pretest.

Figura 9.5
El diseño de cuatro grupos de Solomon



2. Se manifestarán menos prejuicios en el postest del grupo 1 que en el postest del grupo 2.

3. El postest del grupo 3 debe mostrar menos prejuicios que el pretest del grupo 2.

4. El postest del grupo 3 debe mostrar menos prejuicios que el postest del grupo 4.

Observe que los resultados (3) y (4) descartan cualquier interacción entre la prueba y el estímulo. Además, recuerde que estas comparaciones sólo tienen sentido si los sujetos se asignaron al azar a los grupos, con lo que éstos inician con el mismo grado de prejuicios aunque este grado anterior al experimento se mida nada más en los grupos 1 y 2.

Hay un beneficio adicional en este diseño de investigación, como puntualizan los autores. El diseño de cuatro grupos de Solomon no sólo descarta las interacciones entre prueba y estímulo, sino que también arroja datos para comparaciones que revelarán el alcance de tal interacción cuando ocurra en el diseño experimental clásico. Este conocimiento permite al investigador repasar y juzgar el valor de cualesquiera investigaciones anteriores en las que hubiera aplicado este diseño más simple.

El último diseño experimental que quiero mencionar es lo que Campbell y Stanley (1963:25-26)

denominan el **diseño de grupo control con sólo postest**; consta de la segunda mitad —grupos 3 y 4— del diseño de Solomon. Los autores argumentan en forma convincente que, con la aleatorización adecuada, sólo se necesitan los grupos 3 y 4 para un verdadero experimento que controle los problemas de invalidez interna, así como la interacción entre prueba y estímulo. Con la asignación aleatoria de los grupos experimental y control (que distingue a este diseño de la comparación de grupo estático que ya estudiamos), los sujetos serán comparables en cuanto a la variable dependiente —lo bastante comparables para satisfacer las pruebas estadísticas convencionales con que se evalúan los resultados—, así que no es necesario medirlos. En realidad, Campbell y Stanley postulan que en estas circunstancias lo único que justifica el pretest es la tradición. Simplemente, los investigadores se acostumbraron a realizar pretest y se sienten más seguros con los diseños que las incluyen. Sin embargo, tenga en claro que esto se aplica sólo a los experimentos en los que los sujetos se asignaron aleatoriamente a los grupos experimental y control, puesto que eso es lo que justifica la suposición de que los grupos sean equivalentes, sin tener que medirlos para verificarlo.

Confo en que esta exposición le haya dado una idea de las complejidades del diseño experimental, sus problemas y algunas de sus soluciones. Desde luego, hay muchos otros diseños experimentales posibles. Algunos comprenden más de un estímulo y combinaciones de estímulos; otros, varias pruebas sucesivas de la variable dependiente y la presentación del estímulo en diferentes momentos a grupos diversos. Si a usted le interesa profundizar en el tema, consiga el libro de Campbell y Stanley, puesto que las variaciones rebasan el alcance de nuestro texto.

Ejemplo de experimentos

En las ciencias sociales se acude a los experimentos para estudiar una gran variedad de temas. Algunos experimentos se realizan en situaciones de laboratorio; otros ocurren fuera, en el "mundo real". La siguiente exposición le permitirá dar un vistazo de ambos al tiempo que nos concentramos en un solo tema.

En *Pygmalion*, la entrañable obra de George Bernard Shaw, base de la exitosa comedia musical de Broadway, *Mi bella dama*, Eliza Doolittle habla de los poderes que tienen los demás para determinar nuestra identidad social. Distingue el trato que le da su tutor, el profesor Higgins, y el amigo de éste, el coronel Pickering, de la siguiente manera:

¿Sabe? En realidad, además de las cosas que puede una escoger (los vestidos, la manera adecuada de hablar y todo eso), la diferencia entre una dama y una vendedora de flores no es cómo se comporta, sino cómo la tratan. Siempre será una vendedora de flores para el profesor Higgins, porque siempre me trata como vendedora de flores y siempre lo hará. Pero sé que para usted puedo ser una dama, porque usted siempre me trata como dama y siempre lo hará.

El sentimiento que expresa Eliza es básico en las ciencias sociales, y lo han abordado más formalmente sociólogos como Charles Horton Cooley ("el espejo del yo") y George Herbert Mead ("el otro generalizado"). El punto fundamental es que lo que pensamos que somos, nuestro *concepto propio*, y la manera en que nos comportamos está en función, en buena medida, de la manera en que nos ven y nos tratan los demás. Y, relacionado con esto, la forma en que los demás nos perciben está muy

condicionada por lo que esperan por anticipado de nosotros. Por ejemplo, si les dijeron que somos estúpidos, es probable que nos vean así, y acaso nosotros lleguemos a vernos de ese modo y actuemos estúpidamente.

Si usted ya empezó a pensar que estas ideas se prestan a las mil maravillas para experimentos controlados, tiene toda la razón. Más aún, este tema se llama el *efecto Pygmalión*.

En uno de los experimentos más conocidos sobre el tema, Robert Rosenthal y Leonore Jacobson (1968) aplicaron lo que denominaron "Prueba de Harvard de adquisición modulada" a los alumnos de la escuela West Coast. Posteriormente, se reunieron con los maestros de los estudiantes y les mostraron los resultados de la prueba. En particular, Rosenthal y Jacobson, basados en los resultados de la prueba, identificaron a ciertos estudiantes como proclives a mostrar un aceleramiento repentino de sus capacidades académicas durante el año siguiente.

Después, cuando se compararon las puntuaciones en las pruebas de inteligencia, las predicciones de los investigadores resultaron correctas. Los estudiantes identificados como "acelerados" excedieron con mucho a sus condiscípulos durante el año siguiente, lo que indicaría que la prueba de pronóstico fue muy acertada. Pero la prueba era un engaño. Los investigadores hicieron sus predicciones al azar entre alumnos buenos y malos, y lo que dijeron a los maestros no fue producto en absoluto de las puntuaciones de los estudiantes en la prueba. Los progresos que alcanzaron los "acelerados" fueron únicamente resultado de que los maestros esperaron sus mejoras y les prestaron más atención a estos estudiantes, los estimularon y los recompensaron por sus logros. Observe la similitud entre esta situación y el efecto Hawthorne, que citamos antes.

El estudio de Rosenthal y Jacobson llamó mucho la atención, tanto popular como científica. Algunos experimentos posteriores se han concentrado en aspectos concretos de lo que ha llegado a conocerse como el *proceso de atribución* o el *modelo de comunicación de expectativas*. Estas investigaciones, realizadas sobre todo por psicólogos, comparables con los estudios emprendidos por sociólogos, adoptan una postura ligeramente distinta y suelen reunirse bajo el título de *teoría de los estados de expectativa*. Los primeros estudios se concentran en las situaciones en que las expectativas

de un individuo dominante influyen en el desempeño de los subordinados como en el caso de los maestros y los estudiantes, o bien de un jefe y sus empleados. Por su parte, la investigación sociológica ha tendido a concentrarse más en la función de las expectativas entre iguales en grupos pequeños dedicados a determinados cometidos. Por ejemplo, ¿cómo se evalúan los miembros de un jurado cuando se encuentran y cuál es el efecto de estas evaluaciones iniciales en sus relaciones posteriores?

El siguiente es un ejemplo de un experimento realizado para investigar la forma en que las impresiones que tenemos de nuestras capacidades y de los demás influyen en nuestra disposición a aceptar las ideas de otro. Martha Foschi, G. Keith Warringer y Stephen Hart (1985) se interesaban en particular en el papel que juegan los "criterios" al respecto.

En términos generales, entendemos por "criterios" cuán bien o mal tiene que desempeñarse una persona para que se le atribuya o niegue una capacidad. En nuestra opinión, los criterios son una variable clave que influye en la transformación y elaboración de las evaluaciones y en las expectativas que se generan. Por ejemplo, dependiendo de los criterios considerados, puede tomarse el mismo grado de éxito como una gran realización o descartado por insignificante.

(1985:108-109)

Para comenzar a examinar la función de los criterios, los investigadores diseñaron un experimento que comprendía cuatro grupos experimentales y un control. A los sujetos les informaron que el experimento consistía en algo llamado "capacidad de reconocimiento de patrones", y que era una capacidad innata que unos poseían y otros no. Los investigadores les dijeron que trabajarían en parejas sobre problemas de reconocimiento de patrones. Como habrá conjeturado, no existe tal "capacidad".

La primera etapa del experimento fue una "prueba" de las capacidades de cada sujeto en el reconocimiento de patrones. Si usted hubiera sido uno de los sujetos del experimento, le habrían mostrado durante ocho segundos un diseño geométrico seguido por otros dos, parecidos pero no iguales al primero. Su tarea habría sido elegir cuál de los conjuntos subsecuentes tenía el diseño más semejante al primero que vio. Se le pediría que lo repitiera 20 veces, y al cabo una salida de computadora imprimiría su "puntuación". A la mitad de los sujetos se les dijo que habían tenido 14 aciertos; a la otra mitad, que sólo llegaron a seis correc-

tos (sin que importara qué diseños equipararon con cuáles). Dependiendo de su suerte en el sorteo, usted pensaría que lo hizo muy bien o en cambio muy mal. Pero observe que no tendría ningún criterio para evaluar su desempeño (ya que cuatro aciertos podrían considerarse un gran logro).

Asimismo, al tiempo que recibiera su calificación, le darían también la "puntuación de su pareja", aunque tanto las "parejas" como las "puntuaciones" eran ficciones computarizadas (se les dijo a los sujetos que se comunicarían con sus parejas mediante terminales de computadora y que no se les permitiría verse). Si se le hubiera asignado una puntuación de 14, se le habría dicho que su pareja obtuvo seis; de dársele seis a usted, le habrían informado que su pareja consiguió 14.

Este procedimiento significaba que usted pasara a la fase de trabajo en equipo del experimento creyendo o bien que (1) se desempeñó mejor que su pareja, o bien que (2) se desempeñó peor que su pareja. Esto constituyó parte del "criterio" con que operaría durante el experimento. Además, a la mitad de cada grupo se le informó que una puntuación entre 12 y 20 mostraba que el sujeto definitivamente poseía la capacidad de reconocer patrones. A los otros sujetos se les dijo que una calificación de 14 en realidad no era lo bastante alta para probar nada en definitiva. Así, usted habría quedado con una de las siguientes convicciones:

1. Que usted es definitivamente mejor que su pareja para reconocer patrones.
2. Que usted es posiblemente mejor que su pareja.
3. Que usted es posiblemente peor que su pareja.
4. Que usted es definitivamente peor que su pareja.

Al grupo control del experimento no se le dijo nada sobre sus capacidades o de las de sus parejas. En otras palabras, no tenían expectativas.

La última etapa del experimento consistió en poner a trabajar a los "equipos". Como antes, a cada sujeto se le mostraría un diseño inicial seguido de un par de comparación para que escogiera. Pero esta vez, cuando usted anotara su elección, se le diría lo que respondió su pareja y se le pediría que escogiera de nuevo. En su elección final, usted podría aferrarse a su opción inicial o bien cambiarla. Desde luego, la elección de la "pareja" era una creación de la computadora y, como se imaginó, con frecuencia provocaba desacuerdos en el equipo: de hecho, en 16 de cada 20 veces.

La variable dependiente del experimento era el grado al que los sujetos cambiarían sus elecciones para coincidir con las de sus parejas. Los investigadores plantearon la hipótesis de que "el grupo definitivamente mejor cambiaría con menos frecuencia, seguido del grupo probablemente mejor, el grupo control, el grupo probablemente peor y el grupo definitivamente peor, que sería el que cambiaría más a menudo".

A continuación anotamos las veces que los sujetos de los cinco grupos cambiaron sus respuestas. Tenga presente que cada uno tuvo 16 oportunidades de hacerlo. Estos datos indican que todas las expectativas de los investigadores fueron acertadas, con excepción de la comparación entre probablemente peor y definitivamente peor. El último grupo fue, en efecto, el más proclive a cambiar, pero la diferencia es demasiado pequeña para tomarla como una confirmación de la hipótesis (en el capítulo 17 estudiaremos las pruebas estadísticas que permiten tomar estas decisiones).

Grupo	Número promedio de cambios
Definitivamente mejor	5.05
Posiblemente mejor	6.23
Grupo control	7.95
Posiblemente peor	9.23
Definitivamente peor	9.28

En análisis más detallados se descubrió que se mantiene la misma pauta básica tanto en hombres como en mujeres, aunque resultó un poco más clara con éstas que con aquéllos. Éstos son los datos:

	Número promedio de cambios	
	Mujeres	Hombres
Definitivamente mejor	4.50	5.66
Posiblemente mejor	6.34	6.10
Grupo control	7.68	8.34
Posiblemente peor	9.36	9.09
Definitivamente peor	10.00	8.70

Como el alcance de esta clase de investigaciones parece extremadamente restringido, se podría cuestionar su importancia para cualquier tema. Sin embargo, como parte de una empresa de investigación mayor, estos estudios añaden piezas concretas a nuestra comprensión de procesos sociales más generales.

Vale la pena dedicar un minuto o dos a considerar algunas de las situaciones cotidianas en las que los "estados de expectación" tendrían consecuencias muy reales e importantes. Ya cité el caso de las reuniones a deliberar de un jurado. ¿Qué piensa de todas las formas de prejuicios y discriminación? O también reflexione en el modo en que los estados de expectación aparecen en las entrevistas de trabajo, o cuando conoció a los padres de su media naranja. Medítelo y descubrirá otras situaciones en las que estos conceptos de laboratorio se aplican en la vida real.

Grupos focales o de interés

En ocasiones se llevan sujetos al laboratorio para exploraciones menos rigurosas que las que caracterizan a los experimentos controlados. Particularmente en el campo de la investigación de mercado, los científicos sociales reúnen grupos focales para poner a prueba diversos aspectos de algún asunto. Imagine que tiene el propósito de introducir un producto nuevo; por ejemplo, una computadora. De hecho, supongamos que usted inventó una nueva computadora que no sólo maneja procesador de texto, hojas de cálculo, análisis de datos, etc., sino que también contiene una máquina de fax, sintonizador de radio AM/FM y televisión, reproductora de discos compactos, unidad doble para casetes, horno de microondas, enjuague para dentaduras postizas y cafetera. Para resaltar sus características computacionales y de percolado de café, usted piensa llamarla *Computadora*. Calcula que la nueva computadora se venderá en unos 28 000 dólares, y desea saber si la gente querrá comprarla. Sus esperanzas se verán bien cumplidas con un grupo focal.

En un grupo focal, se reúnen en una sala de 12 a 15 personas para participar en una discusión guiada sobre algún tema; en este caso, la posibilidad de aceptación y venta de la *Computadora*. Estos sujetos se eligen de acuerdo con su pertinencia para el tema que se estudia. Dado el costo proyectado de la *Computadora*, quizá los participantes de su grupo focal se limiten a los sectores de ingresos superiores. Podrían pesar otras consideraciones similares en la selección, aunque no es probable que los participantes sean elegidos con métodos rigurosos de muestreo probabilístico, lo que quiere decir que no representarán estadísticamente a ninguna población significativa. Sin embargo, el propósito

del estudio es explorar, más que describir o explicar en ningún sentido definitivo.

Habitualmente, en los estudios se convoca más de un grupo focal, porque se corre el serio peligro de que un solo grupo de 7 o 12 personas sea muy poco característico para alcanzar conocimientos generalizables.

Richard Krueger indica cinco ventajas de los grupos focales:

1. la técnica es un medio de investigación socialmente orientado que capta datos reales en un ambiente social
2. es flexible
3. su validez aparente es elevada
4. arroja resultados rápidos, y
5. es barata.

(1988:47)

La dinámica interna de los grupos focales suele revelar aspectos del tema que acaso el investigador no anticipó y que no habrían surgido en entrevistas individuales. Por ejemplo, en una conversación marginal, quizá dos participantes bromean sobre el riesgo de borrar una letra del nombre del producto. Este apunte nos ahorraría una gran vergüenza más adelante. Una investigación similar le evitó a Chevrolet un intento fracasado de comercializar su modelo Nova en los países de habla española, por el equivoco de pronunciación: "no va".

Krueger también señala algunas desventajas del método:

1. los grupos focales confieren al investigador menos control que las entrevistas individuales
2. es difícil analizar los datos
3. los moderadores requieren habilidades especiales
4. las diferencias entre los grupos pueden ser problemáticas
5. es difícil armar los grupos, y
6. la discusión debe realizarse en un ambiente propicio.

(1988:44-45)

William Gamson (1992) ha examinado con grupos focales la forma en que los ciudadanos estadounidenses encuadran sus opiniones sobre los temas políticos. Luego de escoger cuatro temas —la acción afirmativa, la energía nuclear, las industrias problemáticas y el conflicto árabe-israelí—, Gamson emprendió un análisis de contenidos de prensa para formarse una idea del contexto de los me-

dios de comunicación en que reflexionamos y hablamos sobre la política. Entonces, convocó a los grupos focales para realizar observaciones de primera mano de la forma en que las personas discuten estos temas con sus amistades.

Además, David Morgan (1993) afirmó que los grupos focales son un medio excelente para generar reactivos de cuestionarios de una encuesta subsecuente.

En enero de 1996 se puso en práctica una variación atractiva de la idea de los grupos focales en la Universidad de Texas en Austin. Los organizadores eligieron una muestra aleatoria nacional de 459 "delegados" para que asistieran a una Convención de Temas Nacionales. Durante una semana, los asistentes escucharon las declaraciones de varias figuras políticas sobre diversos temas. Tuvieron la posibilidad de formular preguntas al vicepresidente Al Gore y a los candidatos presidenciales republicanos Phil Gramm, Richard Lugar, Steve Forbes y Lamar Alexander (Beniger, 1996).

Los investigadores que organizaron la convención se interesaban particularmente en medir las opiniones al comienzo y al final para ver si una experiencia política educativa intensiva cambiaría los puntos de vista de los votantes. Por ejemplo, alrededor de dos tercios creían que había ocurrido una descomposición de la familia tradicional en Estados Unidos, y las deliberaciones no alteraron esta proporción. Quienes atribuyeron la descomposición a las presiones económicas aumentaron de 36 a 51 por ciento, mientras que los que decían que obedecía a la falta de valores se redujeron de 58 a 48 por ciento. En los asuntos externos, el porcentaje que coincidía con la declaración "Estados Unidos debe mantener su cooperación militar con otras naciones para enfrentar los puntos conflictivos del mundo" aumentó de 72 a 82 por ciento.

El porcentaje de quienes contestaban "no sé" disminuyó consistentemente y se registró un cambio definitivo en la respuesta a la afirmación "mis opiniones políticas merecen ser escuchadas".

	Antes	Después	Diferencia
Muy de acuerdo	41.0%	68.1%	127.1
Ligeramente de acuerdo	43.4	25.8	217.6
Ligeramente en desacuerdo	8.4	3.6	24.8
Muy en desacuerdo	4.1	1.4	22.7
No sé	3.2	1.1	22.1

En tanto que los experimentos de laboratorio suelen recibir críticas por ser demasiado ajenos a la vida cotidiana, en los grupos focales se hace un

esfuerzo por simular las tertulias comunes en que la gente discute algún tema. Como veremos en seguida, algunos experimentos se llevan a cabo en medio de la vida real y evidencian la importancia de la lógica experimental para nuestra convivencia en la sociedad.

Experimentos "naturales"

Aunque por costumbre equiparemos los términos *experimento* y *experimento de laboratorio*, muchos experimentos sociales científicos ocurren fuera de los ambientes controlados, muchas veces en el curso de los acontecimientos diarios normales. En algunas ocasiones, la naturaleza diseña y ejecuta experimentos que nosotros observamos y analizamos; en otras, quienes toman las decisiones sociales y políticas cumplen con tal función natural.

Por ejemplo, imaginemos que un huracán afectó cierto pueblo. Algunos habitantes sufrieron graves perjuicios económicos y otros escaparon relativamente indemnes. Así, nos preguntaríamos cuáles son las consecuencias en la conducta de padecer un desastre natural. ¿Es más probable que quienes padecieron más tomen precauciones contra desastres futuros que los que no sufrieron tanto? Para dar con las respuestas podríamos entrevistar a los habitantes algún tiempo después del huracán. Les preguntaríamos sobre las precauciones que tomaron antes del fenómeno y las que toman actualmente, y compararíamos a los que más padecieron con quienes sufrieron más bien poco. De esta manera aprovecharíamos un experimento natural que nunca habríamos podido organizar nosotros, aunque fuésemos tan perversos para quererlo.

Un ejemplo semejante procede de los anales de la investigación social de la segunda Guerra Mundial. Cuando ésta terminó, los investigadores emprendieron encuestas retrospectivas sobre la moral en tiempos de guerra entre los civiles de varias ciudades alemanas. Compararon los informes de la moral entre los habitantes de las ciudades más bombardeadas y de las menos agredidas (los bombardeos no redujeron la moral).

Como el investigador debe tomar las cosas como vienen, los experimentos naturales suscitan muchos de los problemas de validez que ya revisamos.

Así, cuando Stanislav Kasl, Rupert Chisolm y Brenda Eskenasi (1981) decidieron estudiar el impacto del accidente nuclear en Three Mile Island (TMI) en los trabajadores de la planta, tuvieron que ser especialmente cuidadosos al diseñar la investigación:

La investigación de desastres es por fuerza oportunista, cuasiexperimental y a posteriori. En la terminología del clásico análisis de Campbell y Stanley sobre los diseños de investigación, nuestro estudio se encuentra en la categoría de "comparación de grupo estático", considerado uno de los diseños más endebles. Sin embargo, las debilidades son potenciales, y su presencia real depende de las circunstancias peculiares de cada estudio.

(1981:474)

La base del estudio fue una encuesta de quienes estaban trabajando en Three Mile Island el 28 de marzo de 1978, cuando falló el sistema de enfriamiento del reactor número 2 y comenzó a fundirse el núcleo de uranio. La encuesta se realizó entre cinco y seis meses después del accidente. Entre otras cosas, el cuestionario midió las actitudes de los empleados hacia el trabajo en plantas de energía nuclear. De haber medido sólo las actitudes de los trabajadores de TMI después de la emergencia, los investigadores no habrían tenido idea de si sus opiniones se modificaron como resultado. Pero mejoraron su diseño de estudio al elegir una planta nuclear cercana y al parecer comparable, abreviada PB, y entrevistaron a sus trabajadores como grupo control. A esto obedece la referencia a la comparación de grupo estático.

Incluso con un grupo experimental y uno control, los autores estaban al tanto de los posibles problemas de su diseño; en particular, que se basaba en la idea de que los dos grupos de trabajadores eran equivalentes, salvo por el hecho único del accidente. Los investigadores habrían podido asumir si hubieran estado en posición de asignar al azar a los trabajadores de las dos plantas, pero desde luego no era el caso. En cambio, tuvieron que comparar las características de los dos grupos e inferir que eran equiparables. A fin de cuentas, los investigadores concluyeron que los dos grupos eran muy parecidos, y que la planta donde trabajaban no era más que una función de su lugar de residencia.

Aun concediendo que los dos conjuntos de trabajadores fueran equivalentes, los investigadores enfrentaban otro problema de comparabilidad, a saber, que no pudieron localizar a todos los trabajadores que estaban en servicio en TMI en el momento del accidente. Los investigadores comentan la dificultad como sigue:

Un problema especial de desgaste del estudio fue la posibilidad de que algunos de los sujetos de TMI, que no respondieron por no haber sido localizados, lo que no sucedió con los sujetos de PB, hubieran abandonado para siempre la región por causa del accidente. Este sesgo por desgaste, con toda probabilidad, atenúa el cálculo estimado del impacto. Con las pruebas de números telefónicos desconectados o "fuera de servicio", estimamos que este sesgo era despreciable (uno por ciento).

(Kasl et al., 1981:475)

El ejemplo de TMI señala tanto los problemas peculiares de los experimentos naturales como la posibilidad de considerar tales dificultades. En general, la investigación social requiere ingenio y perspicacia, pero los experimentos naturales exigen más que el promedio.

Al comienzo del capítulo presentamos un ejemplo hipotético de la proyección de una película sobre la historia de los negros para reducir los prejuicios. Sandra Ball-Rokeach, Joel Grube y Milton Rokeach (1981) pudieron abordar el tema en la vida real mediante un experimento natural. En 1977, se proyectó la dramatización televisiva de *Raíces*, de Alex Haley, por la cadena ABC en ocho noches consecutivas. Sumó las mayores audiencias de la historia de la televisión hasta ese momento. Ball-Rokeach y sus colaboradores querían saber si *Raíces* cambió las actitudes de los estadounidenses blancos hacia los negros. Su oportunidad surgió en 1979, cuando se televisó una continuación (*Raíces: la generación siguiente*). Aunque habría sido bueno desde el punto de vista de los investigadores asignar muestras aleatorias de estadounidenses que vieron o no el programa, no fue posible. En cambio, los investigadores eligieron cuatro muestras del estado de Washington, les enviaron por correo los cuestionarios y midieron sus actitudes hacia los negros. Al terminar el último episodio, se les llamó y se les preguntó si habían visto los episodios, y cuántos vieron. A continua-

ción se enviaron cuestionarios a los que respondieron y se midieron de nuevo sus actitudes hacia los afroestadounidenses.

Al comparar las actitudes anteriores y posteriores tanto de los que presenciaron la emisión como de quienes no lo hicieron, los investigadores llegaron a varias conclusiones. Por ejemplo, descubrieron que las personas que ya sostenían posturas igualitarias tendieron más a ver el programa que quienes tenían más prejuicios contra los negros: un fenómeno de autoselección. Más aún, la comparación de las actitudes anteriores y posteriores de quienes vieron el programa indicó que éste en sí tuvo poco efecto o ninguno. Quienes lo presenciaron no fueron más igualitarios después que antes.

Este ejemplo anticipa la materia del capítulo 14, la *investigación evaluadora*, que puede considerarse una clase de experimento natural. Como veremos, consiste en llevar al campo la lógica de la investigación y evaluar los efectos de los estímulos en la vida real. Puesto que es una forma cada vez más importante de investigación social, le dedicaremos un capítulo entero.

Ventajas y desventajas del método experimental

La principal ventaja de los experimentos controlados radica en el aislamiento de la variable experimental y en su efecto con el paso del tiempo. Esto se aprecia más claramente en el contexto del modelo experimental básico. Al principio del experimento se encuentra que un grupo de sujetos posee cierta característica; luego de la presentación de estímulo, se descubre que tienen una característica distinta. En la medida en que los sujetos hayan experimentado o no otros estímulos, podemos concluir que el cambio de características es atribuible al estímulo experimental.

Más aún, puesto que cada experimento tiene un alcance limitado y requiere relativamente poco dinero, tiempo y sujetos, a menudo podemos repetirlo varias veces con grupos diferentes de sujetos (desde luego, no siempre ocurre así, pero es más fácil repetir experimentos que, digamos, encuestas). Como en todas las otras formas de investigación científica, la repetición de los hallazgos fortalece nuestra confianza en su validez y generalización.

La principal desventaja de los experimentos de laboratorio radica en su artificialidad. Los procesos sociales que se desenvuelven en el ambiente del laboratorio no suceden necesariamente en los medios sociales más naturales. Por ejemplo, un filme de la historia de los negros bien podría reducir los prejuicios de los sujetos de un grupo experimental; sin embargo, esto no significa que el mismo filme proyectado en los cines de barrio de todo el país reducirá por fuerza los prejuicios del público en general. Desde luego, la artificialidad no es tanto un problema de los experimentos naturales como de los efectuados en laboratorio.

Al repasar las fuentes de invalidez interna y externa que citan Campbell y Stanley, vimos que es posible idear diseños experimentales que controlen en forma lógica tales problemas. Esta posibilidad señala una de las grandes ventajas de los experimentos: prestarse a un rigor lógico que es mucho más difícil de conseguir en otros modos de observación.

Puntos principales

- Los experimentos son un vehículo excelente para la prueba controlada de los procesos causales.
- El experimento clásico estima el efecto de un estímulo sobre alguna variable dependiente mediante pretest y posttest de los grupos experimental y control.
- En general, es menos importante que un grupo experimental de sujetos sea representativo de una población mayor que el hecho de que éste y el grupo control sean similares.
- La aleatorización suele ser el método preferido para lograr la comparabilidad de los grupos experimental y control.
- Campbell y Stanley describen tres formas de preexperimentos: el estudio de caso de una medición, el diseño pretest-posttest de un grupo y la comparación de grupo estático.
- Hay 12 fuentes de invalidez interna en el diseño experimental:
 1. Historia
 2. Maduración
 3. Administración de pruebas
 4. Instrumentación
 5. Regresión estadística
 6. Sesgos en la selección
 7. Mortalidad experimental
 8. Orden causal temporal
 9. Difusión o imitación de tratamientos
 10. Compensación
 11. Rivalidad compensatoria
 12. Desmoralización
- El experimento clásico con asignación aleatoria de sujetos nos precave de estas fuentes de invalidez interna.
- Los experimentos también enfrentan problemas de invalidez externa: sus hallazgos pueden no reflejar la vida real.
- La interacción de prueba y estímulo es un ejemplo de invalidez externa; el experimento clásico no nos libra de ella.
- El diseño de cuatro grupos de Solomon y otras variaciones del experimento clásico pueden solventar la invalidez externa.
- Campbell y Stanley proponen que, con una aleatorización conveniente en la asignación de los sujetos a los grupos experimental y control, no hay necesidad de realizar pretest en los experimentos.
- Los experimentos naturales ocurren en el curso de la vida social en el mundo real. Además, los investigadores sociales pueden ponerlos en práctica más o menos de la misma manera con que diseñan y efectúan experimentos de laboratorio.

Preguntas y ejercicios de repaso

1. Elija seis de las 12 fuentes de invalidez interna citadas en el libro y mencione ejemplos (que no hayamos visto) para ilustrarlas.
2. Piense en un desastre natural que haya atestiguado o del que haya leído. Enmarque una pregunta de investigación que pueda estudiarse tratando al desastre como experimento natural. Esboce en dos o tres párrafos la manera en que realizaría el experimento.
3. En este capítulo consideramos brevemente el problema del "efecto del placebo". Busque en internet un estudio en el que este efecto figure

de manera importante. Escriba un informe corto sobre el estudio que incluya la fuente de su información. (Sugerencia: Si lo desea, realice una búsqueda con la palabra *placebo*.)

4. Supongamos que le pidieron que hiciera una investigación para determinar la aceptabilidad de una nueva computadora de mano. Planee una sesión de grupo de interés que arroje información útil para este propósito.

Proyecto de continuidad

Piense en un estímulo experimental que pudiera influir en las opiniones de los estudiantes sobre la igualdad sexual. Describa un experimento que ponga a prueba tal estímulo.

Lecturas adicionales

Campbell, Donald, y Julian Stanley, *Experimental and Quasi-Experimental Designs for Research*, Chicago, Rand McNally, 1963. Excelente análisis de la lógica y los métodos de la experimentación en la investigación social. El libro es especialmente útil por su aplicación de la lógica de los experimentos a otros métodos de investigación social. Si bien es un tanto antiguo, se trata de una obra que ha alcanzado el estatus de clásica y aún se cita con frecuencia.

Cook, Thomas D., y Donald T. Campbell, *Quasi-Experimentation: Design and Analysis Issues for*

Field Settings, Chicago, Rand McNally, 1979. Versión ampliada y puesta al día del libro de Campbell y Stanley.

Jones, Stephen R. G., "Worker Independence and Output: The Hawthorne Studies Reevaluated", en *American Sociological Review* 55 (abril de 1990): 176-190. El artículo revisa estos artículos clásicos y cuestiona la interpretación tradicional (que es la que se presenta en este capítulo).

Martin, David W., *Doing Psychology Experiments*, Monterey, Brooks/Cole, 1991. Explicaciones minuciosas de la lógica de los métodos de investigación, a menudo en tono humorístico. El libro destaca las ideas de particular importancia para el investigador principiante, como conseguir un tema para un experimento o revisar la bibliografía.

Morgan, David L. (comp.), *Successful Focus Groups: Advancing the State of the Art*, Newbury Park, Cal., Sage, 1993. Esta colección de artículos sobre las aplicaciones de los grupos focales toca muchos aspectos que normalmente no se consideran.

Ray, William, y Richard Ravizza, *Methods toward a Science of Behavior and Experience*, Belmont, Cal., Wadsworth, 1993. Examen general de los métodos de investigación de las ciencias sociales, con el acento puesto en la experimentación. Este libro es especialmente sólido en la filosofía de la ciencia.

