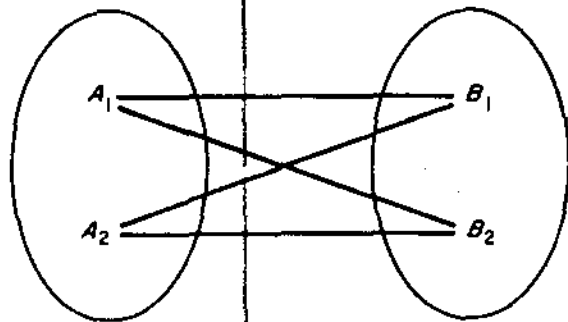


Diseños generales de investigación

El diseño es la disciplina de los datos. El objetivo implícito de cualquier diseño es imponer restricciones controladas a las observaciones de fenómenos naturales.

Efectivamente, el diseño le ordena al investigador: haz esto o aquello; no hagas esto ni aquello; ten cuidado con esto; ignora aquello, etc. Para el investigador viene a ser lo mismo que el anteproyecto para el ingeniero y el arquitecto. Si su estructura está concebida de manera deficiente, el producto final será defectuoso. Si por lo menos la estructura está bien concebida, el producto tiene mayor probabilidad de que se le considere valioso desde el punto de vista científico. En este capítulo nuestra gran preocupación es presentar siete u ocho "buenos" diseños básicos.

Además de ello, trataremos sobre algunos fundamentos conceptuales de la investigación y de dos o tres problemas relacionados con el diseño; por ejemplo, los motivos por los que se recurre a los grupos de control y las ventajas y desventajas del apareamiento.



Así pues, los pares ordenados son: A_1B_1 , A_1B_2 , A_2B_1 , A_2B_2 . Esto es una relación pues tenemos un conjunto de pares ordenados. Además es una partición cruzada. El lector deberá observar las figuras 4-7 y 4-8 del capítulo 4, ya que le ayudarán a clasificar estos conceptos y a comprender la aplicación del producto cartesiano y las ideas de relación al diseño de investigación. Por ejemplo A_1 y A_2 pueden ser los aspectos de cualquier variable independiente: grupo

Fundamentos conceptuales del diseño de investigación

El fundamento conceptual para entender el diseño se expuso en los capítulos 4 y 5; en ellos se definieron y explicaron los conjuntos y las relaciones. Recuerdese que una *relación* es un conjunto de pares ordenados. (También puede ser un conjunto de triples, cuádruples, etc.) Recuerdese asimismo que el *producto cartesiano* son todos los posibles pares ordenados de dos conjuntos. La *partición* desglosa un conjunto universal U en subconjuntos que son *disjuntos* y *exhaustivos*. La *partición cruzada* es una nueva segmentación que se obtiene al dividir el conjunto U y formar todos los subconjuntos de la forma $A \cap B$. Estas definiciones fueron explicadas ampliamente en los capítulos 5 y 6. Ahora, las aplicaremos a las ideas de diseño y análisis.

Tomemos dos conjuntos, A y B , divididos en A_1 y A_2 , B_1 y B_2 . El producto cartesiano de dos conjuntos es:

experimental, grupo de control, dos métodos didácticos un varón y una mujer, etcétera.

Un *diseño* es algún subconjunto del producto cartesiano de las variables independientes y de la variable dependiente. Es posible parrear cada medida de la variable dependiente, que representaremos con Y en nuestro texto, con algún aspecto o partición de una variable independiente. Los casos más sencillos que pueden presentarse ocurren con una variable inde-

pendiente y otra dependiente. En el capítulo 10, una variable independiente (A) y una variable dependiente (B) fueron divididas en $\{A_1, A_2\}$ y $\{B_1, B_2\}$; luego se hizo una partición para formar la ya conocida división cruzada 2×2 , con frecuencias o porcentajes dentro de las casillas.

Nos concentraremos en particiones cruzadas similares de A y B , pero que tengan medidas continuas en las casillas.

Tomemos a A por separado, valiéndonos del diseño de análisis de varianza de una sola variable. Supongamos tres tratamientos experimentales, A_1 , A_2 y A_3 , y por razón de simplicidad dos puntuaciones de Y en cada casilla. Ello se advierte en la parte izquierda de la figura 19-1, designada por (a). Supongamos que seis sujetos fueron asignados aleatoriamente a los tres tratamientos y que sus puntuaciones al final del tratamiento son las que aparecen en la figura.

El lado derecho de la figura 19-1, designado con (b), muestra la misma idea en forma de pares ordenados o de relación. Los pares ordenados son A_1Y_1 , A_1Y_2 , A_2Y_1 , A_2Y_2 , A_3Y_1 , A_3Y_2 . Por supuesto no es un producto cartesiano, que acoplaría a A_1 con todas las Y , A_2 con todas las Y , y A_3 con todas las Y , para dar un total de $3 \times 6 = 18$ pares. Más bien, la figura 19-1(b) es un subconjunto del producto cartesiano, $A \times B$. Los diseños son subconjuntos de $A \times B$, y tanto el diseño como el problema de investigación definen o especifican cómo se construyen estos subconjuntos. Los del diseño de la figura 19-1 presumiblemente obedecen al problema de investigación.

Cuando hay más de una variable independiente la situación se torna más compleja. Tomemos dos variables independientes, A y B , divididas en $\{A_1, A_2\}$ y $\{B_1, B_2\}$. (1) Ahora debemos tener triples ordenados (o dos conjuntos de pares ordenados): ABY . Examine la figura 19-2. En el lado izquierdo, designado con (a), aparece el análisis factorial 2×2 de diseño de varianza y el ejemplo usado en el capítulo 14 (cuadro 14-3), con las medidas de la variable dependiente, Y , insertadas dentro de las casillas. Es decir, ocho sujetos fueron asignados aleatoriamente a las cuatro casillas. Al final del experimento sus puntuaciones son Y_1, Y_2, \dots, Y_8 . El lado derecho de la figura, identificado por (b), muestra en forma de árbol los triples ordenados ABY . Es evidente que son subconjuntos de $A \times B \times Y$ y que son relaciones. El mismo razonamiento es aplicable a diseños más amplios e intrincados, como el diseño factorial $2 \times 2 \times 3 (ABCY)$ o $4 \times 3 \times 2 \times 2 (ABCDY)$. (En estas designaciones, Y suele omitirse porque está implícita.) Otras clases de diseños pueden conceptualizarse de manera similar, aunque resulta difícil describirlos en forma de árboles. Lo único que desea-

(1) El lector debe evitar confundir esto con el paradigma anterior AB de frecuencia, donde A era la variable independiente y B la variable dependiente.

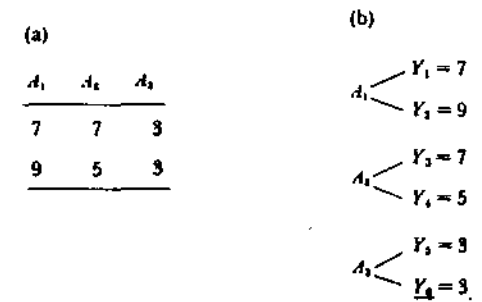


Figura 19-1.

mos es destacar las ideas fundamentales del diseño y no necesitamos dar más ejemplos.

En resumen, el diseño de investigación es algún subconjunto del producto cartesiano de las variables independientes y dependientes. Si se tiene una sola variable independiente, se hacen particiones de esta variable, si hay más de una, se hace una partición cruzada de estas variables independientes. En caso de tres o más, la conceptualización es idéntica, pero difieren las dimensiones; por ejemplo, $A \times B \times C$ y $A \times B \times C \times D$ y las particiones cruzadas que se obtengan de ellas. Siempre que sea posible, conviene contar con diseños "completos" —el diseño completo es una partición cruzada de las variables independientes(2)— y observar las dos condiciones básicas de desunión y exhaustividad. Dicho de otro modo, el diseño no debe incluir el mismo caso (una puntuación del sujeto) en más de una casilla de una partición o una partición cruzada, y todos los casos deben agotarse. Más aún, el mínimo esencial de cualquier diseño es una partición de la variable independiente en dos subconjuntos; por ejemplo, A en A_1 y A_2 . (En este sentido, los diseños 18-1 y 18-2 son incompletos.)

El término "diseños generales" significa que los que se presentan en este capítulo se simbolizan o expresan en su forma más universal y abstracta. Donde se encuentre una simple X , que representa a la variable independiente, deberá entenderse que denota más de una X , o sea de la X se ha hecho una partición en dos o más grupos experimentales. Por ejemplo, el diseño 19-1 que se estudiará a continuación tiene X y $\sim X$, que designan los grupos experimental y de control y que por tanto son una partición de X . Pero ésta puede fraccionarse en varias X , para lo cual el diseño se convierte de un diseño simple con una variable en, digamos, uno factorial. A pesar de ello, el simbolismo básico del diseño 19-1 permanece inalterable. Confiamos que en este y en los siguientes

(2) Hay también diseños "incompletos", pero a los completos se les da más importancia en este libro. Ver a R. Kirk, *Experimental Design: Procedures for the Behavioral Sciences*. Belmont, Calif.: Brooks Cole, 1968, en especial el cuadro 1.4-1, pág. 12.

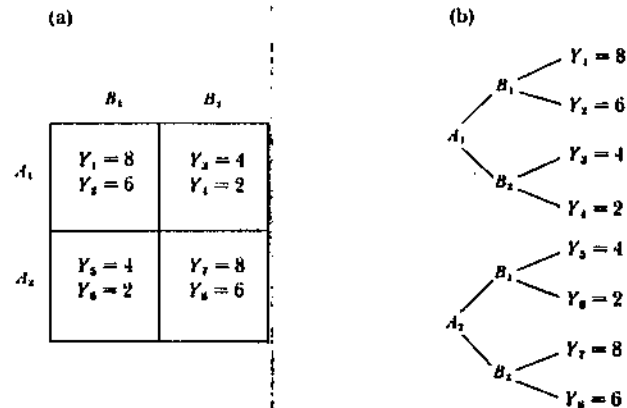


Figura 19-2.

tes capítulos se aclaren las complejidades. Por ahora baste recordar que los diseños son relaciones donde los subconjuntos de pares ordenados son los símbolos de las variables independientes $A_1, A_2, B_1, B_2, X, \sim X$, etc., y las medidas de la variable dependiente, Y . Los diseños completos son productos cartesianos de las variables independientes, y todos los diseños son alguna forma de particiones o particiones cruzadas, sujetas a las reglas de partición.

Debe aclararse un último punto antes de abordar los diseños. Estos surgieron debido a las necesidades de la experimentación; se inventaron para imponer control a las variables experimentales. Esto ya lo vimos al hablar del análisis de varianza. Aunque algunos diseños inspirados en el análisis de varianza han sido adaptados para utilizarlos en la investigación *ex post facto*, su razón fundamental se encuentra en las ideas y situaciones experimentales. Así pues, junto con la noción de que el experimento controlado es el ideal de la investigación científica, hemos de afirmar que el diseño posee una concepción esencialmente experimental. Su aplicación a la investigación *ex post facto* es por lo menos discutible, y algunas veces difícil e inapropiada. No obstante, sostenemos que es mejor contar con un diseño bien concebido que con uno incorrecto o nada, y trataremos de hacer todo lo posible por fortalecer tanto la investigación experimental como la investigación *ex post facto*.

Diseños

A diferencia de los diseños del capítulo anterior, los de esta sección no tendrán equivalentes de tipo *ex post facto* porque estos últimos excluyen la aleatorización. Este problema se expondrá con mayor detalle en un capítulo ulterior.

Diseño 19-1: sujetos asignados aleatoriamente al grupo de control y al grupo experimental

R	X	Y	(Grupo experimental)
	$\sim X$	Y	(Grupo de control)

Este diseño, con los dos grupos anotados arriba y sus variantes con más de dos grupos, son probablemente los "mejores" para los numerosos fines experimentales en el campo de la educación y la psicología. Es evidente la semejanza de su estructura con la del diseño 18-4. La *R* antepuesta al paradigma indica que los sujetos han sido asignados aleatoriamente al grupo experimental (hileras superior) y al grupo de control (hileras inferior), con lo cual caen por tierra las objeciones mencionadas en el capítulo 18. Aunque en teoría todas las variables independientes posibles están bajo control, en la práctica no siempre ocurre así. Si se incluye en el experimento un número suficiente de sujetos que hagan operable el principio de aleatorización, tendremos sin duda un potente control. En otras palabras, los requisitos de validez interna se cumplirán perfectamente.

No podemos decir lo mismo de la validez externa. Casi siempre resulta imposible seleccionar al azar a los sujetos entre una población determinada. Debemos contentarnos, o al menos resignarnos, con una selección no aleatoria. Por tanto, aunque el diseño reúne las condiciones de validez interna, no cabe hacer generalizaciones más allá de las muestras a nuestra disposición, si los sujetos (y, en teoría, las situaciones) no han sido escogidos al azar. Pero el problema no es tan difícil de resolver como parece. Es posible acrecentar considerablemente la generalizabilidad mediante la reproducción de experimentos con y sin variantes. Por ejemplo, si al repetir un

experimento en diferentes momentos y lugares persisten las supuestas relaciones en todos ellos, tendremos mucho más confianza en la validez científica de las relaciones.

Si se extiende a más de dos grupos y si es capaz de responder a las preguntas planteadas, este diseño ofrecerá entonces las siguientes ventajas: 1) tendrá el mejor sistema de control interno teórico para cualquier diseño, con una o dos excepciones en casos especiales; 2) será flexible, porque en teoría es susceptible de extensión a cualquier número de grupos con cualquier número de variables; 3) si se extiende a más de una variable permitirá demostrar varias hipótesis a la vez, y 4) estará dotado de elegancia estadística y probabilística.

No pretendemos afirmar que sea el mejor y definitivo: no es universal ni perfecto. Es claro que en su forma actual no puede verificar los experimentos del cambio con la misma exactitud con que lo hacen los diseños de preprueba y posprueba (pretest y posttest); y en caso de extenderse a más de una variable independiente, las consideraciones de carácter práctico limitan su aplicabilidad en la investigación de la conducta. No obstante, desde el punto de vista de la investigación es el más eficaz por dos razones:

1. Gracias al grupo de control (o varios grupos experimentales) es posible la comparabilidad requerida por la ciencia. Las comparaciones son esenciales en toda investigación científica. El diseño clásico de grupo experimental-grupo de control, en que se insertan ambos (mediante la asignación aleatoria), permite realizarlas con eficacia, lo cual hace de él un logro intelectual de primer orden.
2. La segunda razón de su eficacia es la aleatorización. A fin de validar la idea de grupo testigo-control experimental, el experimentador debe tener cierto grado de seguridad de que sus grupos son aproximadamente (estadísticamente) iguales en cualquier variable relacionada con la variable o variables dependientes. Según vimos, ello no se puede lograr con el apareamiento ni con los juicios intuitivos o experienciales.

A muchos estudiantes les es difícil aceptar que una tabla de números aleatorios sea tan eficaz. Confiamos en que nuestras anteriores demostraciones Monte Carlo los hayan convencido. También confiamos en que nuestras abundantes citas de estudios factoriales hayan confirmado su fe en las extensiones multivariantes de este diseño: $2 \times 2, 3 \times 2, 3 \times 2 \times 4$, y así sucesivamente.

Antes de abordar los demás diseños, hemos de examinar la noción de grupo de control (una de las grandes invenciones de los últimos cien años) y también ciertas extensiones de este diseño. Ambos temas se complementan perfectamente.

El concepto de grupo de control y extensiones del diseño 19-1. Evidentemente la palabra "control" y la expresión "grupo de control" no aparecían en la

literatura científica anterior al siglo XIX.⁽³⁾ La noción de experimentación controlada es mucho más antigua; dice Boring que Pascal la utilizó ya en 1648. Solomon buscó en la literatura psicológica y no pudo hallar un solo caso en que se usase el grupo de control antes de 1901.⁽⁴⁾ Afirma que el diseño con grupo de control hubo de aguardar la aparición de avances estadísticos y a que los psicólogos dominaran las técnicas estadísticas.

Quizá en 1901 se emplearon por primera vez los grupos de control en psicología.⁽⁵⁾ Uno de los dos hombres que realizaron la investigación, E. L. Thorndike, aplicó a la educación las ideas básicas y revolucionarias de esta primera serie de trabajos.⁽⁶⁾ En su gigantesco estudio de 8 564 alumnos que asistían a las escuelas de un gran número de ciudades, los grupos de control eran grupos educacionales independientes. Entre otras comparaciones, equiparó los incrementos de las puntuaciones en una prueba de inteligencia, supuestamente producidos por el estudio de inglés, historia, geometría y latín con los incrementos atribuibles al estudio de inglés, historia, geometría y taller pedagógico. Trató de comparar la influencia del latín y del taller pedagógico e hizo otras comparaciones análogas. Pese a las limitaciones del diseño y del control, los experimentos de Thorndike y los efectuados por otros investigadores a quienes estimuló fueron notables por su penetración. Llegó incluso a censurar a las universidades por rechazar a estudiantes de taquigrafía y mecanografía que no habían cursado latín, pues creía haber demostrado que era similar la influencia de varias asignaturas en la inteligencia. Es interesante señalar que, en su opinión, se requerían enormes números de sujetos: exigió 18 000 casos más. También se dio cuenta, en 1924, de la necesidad de que las muestras fueran elegidas aleatoriamente.⁽⁷⁾

La noción de grupo de control debe ser generalizada. Supongamos que en un experimento de pedagogía hay cuatro grupos experimentales: A_1 es el de

(3) E. Boring, "The Nature and History of Experimental Control," *American Journal of Psychology*, LXVII (1954), 573-589.

(4) R. Solomon, "An Extension of Control Group Design," *American Journal of Psychology*, LXVII (1954), 573-589. Tal vez la noción de grupo de control se haya usado en otros campos, aunque se duda que la idea estuviese bien desarrollada. Solomon (*ibid.*, pág. 175) asegura que el estudio de Peterson and Thurstone sobre las actitudes en 1933 fue la primera tentativa seria por valerse de grupos de control en la valoración de los efectos producidos por los métodos educativos. La expresión "grupo de control" no se encuentra en ninguna parte de la famosa onceava edición (1911) de la *Encyclopaedia Britannica*, a pesar de que en ella se habla del método experimental.

(5) E. Thorndike y R. Woodworth, "The Influence of Improvement in One Mental Function upon the Efficiency of Other Functions," *Psychological Review*, VIII (1901), 247-261, 384-395, 553-564.

(6) E. Thorndike, "Mental Discipline in High School Subjects," *Journal of Educational Psychology*, XV (1924), 1-22, 83-98.

(7) Ver *ibid.*, págs. 93, 97, y 85, donde se tratan los tres puntos mencionados.

refuerzo por cada respuesta; A_2 , el de refuerzo en intervalos regulares; A_3 , el de refuerzo a intervalos fortuitos; A_4 , ausencia de refuerzo. Desde el punto de vista técnico, existen tres grupos experimentales y un grupo de control en el sentido tradicional de la expresión. No obstante, A_4 pudiera ser otro "tratamiento experimental"; tal vez una especie de refuerzo mínimo. Así, en el sentido tradicional no habría un grupo de control, pues esa acepción carece de generalidad. Si se generaliza la noción de control, desaparece la dificultad. Siempre que tengamos más de un grupo experimental y cualquiera de ellos reciba tratamientos distintos, existirá control en el sentido de comparación previamente mencionado. Siempre que se procure hacer los dos grupos sistemáticamente diferentes en una variable dependiente será posible la comparación. En síntesis, la noción tradicional de que el grupo experimental debe ser sometido a un tratamiento que no se da al grupo de control es un caso especial de la regla más general, según la cual los grupos de comparación son esenciales para la validez interna de cualquier investigación científica.

Si suponemos que este razonamiento es correcto podemos elaborar diseños como el siguiente:

R	X_1	Y
	X_2	Y
	X_3	Y
R	X_{1a}	Y
	X_{1b}	Y
	X_{2a}	Y
	X_{2b}	Y
	X_{3a}	Y

Estos diseños se reconocerán más fácilmente si se presentan con una notación distinta en la forma que usamos en la parte 5, como se aprecia en la figura 19-3. El de la izquierda es un simple análisis unidireccional del diseño de varianza y el de la derecha es un diseño factorial de tipo 2×2 . En el diseño de la derecha, X_{1a} pudiera ser el grupo experimental

X_1	X_2	X_3
MEDIDAS DE Y		

y X_{1b} , el grupo de control con X_{2a} y X_{2b} como variable manipulada o variable atributiva dicotómica. Por supuesto, se trata del mismo diseño que aparece en la figura 19-2(a).

Diseño 19-2: grupo experimental-grupo testigo, con sujetos igualados

M	X	Y	(Grupo experimental)
	$\sim X$	Y	(Grupo de control)

Su estructura es igual a la del diseño 19-1; la única diferencia es que, en vez de asignar aleatoriamente los sujetos a los grupos, se les aparea en uno o más atributos. Para que se le considere un diseño "adecuado", la "aleatorización" debe figurar en él, como lo expresa la r al pie de la M (que significa "apareados"). No basta usar sujetos igualados; los miembros de cada par deben asignarse al azar a los dos grupos. Asimismo, lo ideal sería decidir con una técnica aleatoria si un grupo será el experimental o el de control. En uno y otro caso, para tomar una decisión, se puede lanzar al aire una moneda o recurrir a una tabla de números aleatorios, de manera que los números impares representen un grupo y los pares el otro. Es claro que si se dispone de más de dos grupos, habrá que usar un sistema de números aleatorios.

Igual que en el diseño 19-1, es posible, aunque no fácil, usar más de dos grupos. (Ya se expuso antes la dificultad que implica aparear a más de dos grupos.) Sin embargo, hay ocasiones en que el diseño de apareamiento es un elemento intrínseco en la investigación. Cuando los mismos sujetos intervienen en dos o más tratamientos experimentales o cuando están dados en más de un intento, el apareamiento es inherente a la situación. En la investigación pedagógica, si las escuelas o clases son realmente variables —por ejemplo, si participan dos o más y los tratamientos experimentales se aplican en cada una— el diseño 19-2 será la base lógica. Examínese el paradigma de un diseño para escuelas que aparece en la figura 19-4. Se advierte que es fácil estimar la varianza resultante de las diferencias entre escuelas, y esa varianza puede ser substancial.

		X_1	
		X_{1a}	X_{1b}
X_2	X_{2a}	MEDIDAS DE Y	
	X_{2b}		

Figura 19-3.

Ventajas y desventajas del apareamiento

El apareamiento no se limita a los sujetos. Nos será fácil captar este punto si lo concebimos en términos de la varianza. Si ciertos subconjuntos del conjunto total de unidades muestreadas se asemejan más que otros, probablemente habrá varianza a causa de las diferencias entre ellos. El caso extremo se presenta cuando los mismos sujetos intervienen en más de un ensayo, porque un sujeto se parece naturalmente más a sí mismo que a los demás. Los tres grupos experimentales de la figura 19-4 pueden ser los mismos sujetos en diversos ensayos, lo cual conlleva una varianza sistemática: la de diferencias individuales. Si se utilizan prepruebas y pospruebas, huelga decir que también existirá apareamiento. Se sabe que las escuelas se distinguen en importantes

características: difieren las clases, las zonas escolares, los vecindarios, los maestros. Tales diferencias pueden usarse en el estudio, y las varianzas resultantes de su empleo pueden aislarse al incorporar sus fuentes en el diseño. De hecho, si no se les integra en él, pueden confundirse las variables experimentales.(8) Es un error grave desde el punto de vista estadístico y del diseño, contar con una correlación entre grupos, debida al apareamiento, a la repetición de medidas de individuos o a unidades como clases y escuelas, y no aprovecharla.(9)

Diseño 19-3: grupo de control sometido a preprueba y posprueba

(a)	R	Y_1	X	Y_2	(Grupo experimental)
		Y_3	$\sim X$	Y_4	(Grupo de control)
(b)	M_{pr}	Y_1	X	Y_2	(Grupo experimental)
		Y_3	$\sim X$	Y_4	(Grupo de control)

Este diseño ofrece muchas ventajas y es de uso corriente. Su estructura es similar a la del diseño 18-2, con dos notables diferencias: al 18-2 le falta un grupo de control y la aleatorización. El 19-3 se parece a los diseños 19-1 y 19-2, con la diferencia de que se ha introducido la característica de preprueba. Se le emplea comúnmente para estudiar el cambio. Igual que los diseños 19-1 y 19-2, puede ampliarse para abarcar más de dos grupos.

En el diseño 19-3(a), los sujetos se asignan aleatoriamente al grupo experimental (hilera superior) y al grupo de control (hilera inferior) y se les somete a una prueba en la medida de Y , la variable dependiente. El investigador puede así proceder a verificar la igualdad de ambos grupos en Y . Se aplica la manipulación experimental X , y luego se vuelven a medir los grupos en Y . Con pruebas estadísticas se comprueba la diferencia entre ellos. Una peculiaridad interesante y difícil de este diseño lo constituye el carácter de las puntuaciones ordinariamente analizadas: puntuaciones de diferencia o cambio, $Y_2 - Y_1 = D$. No se recomienda el análisis de las puntuaciones de diferencia, a menos que el efecto de la manipulación experimental sea notable. Sin embargo, si se les analiza, deberá aplicarse una prueba t o F . Debido a estas y otras dificultades, fácilmente se podría dedicar un capítulo entero a este diseño. El estudiante encontrará una explicación más amplia en la monografía de Campbell y Stanley (véase la nota 14, págs. 13-24). Nos limitaremos a comentar las ventajas y deficiencias del diseño. Al final de la

exposición se tratarán las dificultades analíticas que acabamos de mencionar.

Probablemente lo más importante es que este diseño supera la gran carencia del diseño 18-2, pues proporciona un grupo de control con el cual establece la comparación y verifica la diferencia, $Y_2 - Y_1$. Con un grupo solamente, jamás sabremos si la

(8) El término confusión o conglobación, a menudo aparece en tratados del diseño estadístico y de la investigación, significa la "mezcla" de varianzas de una o más variables independientes —generalmente extrañas a la finalidad de la investigación— con la variable o variables independientes del problema. De ahí que no pueda afirmarse que la relación descubierta se da entre las variables independientes y la variable dependiente de la investigación, o entre las variables independientes extrañas y la variable dependiente, o bien entre ambas. Señala Underwood que solamente hay un principio básico del diseño: "... diseñe el experimento de modo que los efectos de las variables independientes puedan ser valorados sin ambigüedades". (B. Underwood, *Psychological Research*, Nueva York: Appleton, 1957, pág. 86.) Si esto no es posible por tratarse de un procedimiento difícil, la mayoría de las veces se debe a que las variables independientes han sido confundidas. El término deriva seguramente de la estadística, ciencia en que "la confusión" se practica a veces con toda deliberación.

(9) Cabe subrayar que el apareamiento deliberado no es en general un procedimiento recomendable, y que el apareamiento nunca es un substitutivo de la aleatorización. Recuerde que la correlación que exista entre la variable o variables del apareamiento y la variable dependiente ha de ser substancial (superior a 0.50, por ejemplo) para que sea productiva, y que únicamente es posible controlar a la variable o variables apareadas —y acaso a las variables substancialmente correlacionadas con ellas. Un mejor procedimiento suele ser el análisis de covarianza o algún otro procedimiento de regresión, tema que examinaremos en el capítulo 21 y en los últimos.

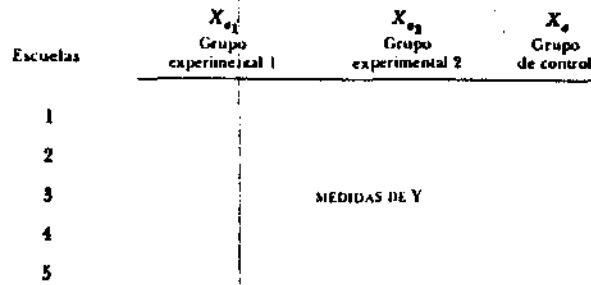


Figura 19-4.

historia, la maduración (o ambas), o la manipulación experimental X produjeron el cambio en Y . Al agregar un grupo de control, la situación se altera radicalmente. Después de todo, si se aparean los grupos (por medio de la aleatorización) los efectos de historia y maduración, en caso de que existan, se manifestarán en ambos grupos. Si aumentan las edades mentales de los niños del grupo experimental, lo mismo debe suceder con la de los niños del grupo de control. En consecuencia, si todavía existe una diferencia entre las medidas de Y en uno y otro grupo, no se deberá a la historia ni a la maduración. Es decir, si algo afecta a los sujetos experimentales entre la preprueba y la posprueba, ese algo deberá afectar por igual a los sujetos del grupo de control. Además, hay que controlar el efecto de la aplicación de las pruebas a las que Campbell llama medidas reactivas. Si la práctica influye en los miembros del grupo experimental, influirá de manera semejante en los del grupo de control. (No obstante, en todo esto hay un punto débil oculto que se comentará más adelante.) Esta es la principal ventaja del diseño cuando está bien planeado y ejecutado.

Por otra parte, los diseños a base de prepruebas y pospruebas ofrecen un aspecto problemático, que merma la validez externa del experimento aunque deje intacta la validez interna. La causa de dicha dificultad es la preprueba, ya que puede ejercer un efecto sensibilizador sobre los sujetos. Por ejemplo, quizás se sientan alterados ante ciertos eventos de su ambiente en los que normalmente no se fijarían. Si la preprueba es una escala de actitudes, es capaz de sensibilizarlos a las cuestiones o problemas mencionados en la escala. Por ello, cuando el tratamiento X se aplica al grupo experimental, acaso los sujetos no sólo respondan a la influencia intencional (la comunicación o cualquier otro método que sirva para modificar las actitudes) sino más bien a una combinación de su mayor sensibilidad ante las preguntas y la manipulación experimental.

Estos efectos de interacción no son obvios a primera vista y constituyen una amenaza contra la vali-

dez externa de los experimentos, por lo cual vale la pena examinarlos un poco más a fondo. Se podría pensar que, como ambos grupos son sometidos a una prueba preliminar, el efecto de ella, en caso de que exista, garantizaría la validez del experimento. Supóngase que no se realizó ninguna prueba previa, esto es, que se usó el diseño 19-2. En igualdad de circunstancias, cabe suponer que se debe a X una diferencia entre el grupo experimental y el grupo de control después de la manipulación experimental de X . No hay motivo para pensar que un grupo es más sensible o alerta que el otro, puesto que ambos afrontan la situación de la prueba después de X . Pero cuando se recurre a una preprueba los resultados difieren. Si bien la preprueba sensibiliza a uno y otro grupo, puede hacer que los sujetos del grupo experimental respondan a X , total o parcialmente, en virtud de la sensibilidad. Así pues, nos hallamos ante una falta de generalizabilidad: se puede generalizar a los grupos que fueron sometidos a la preprueba, no así a los que no fueron sometidos. Claro está que semejante situación molesta al investigador, pues ¿quién quiere generalizar a los grupos que hicieron una preprueba?

Si esta limitación es considerable, ¿por qué decimos que es un buen diseño? Aunque el posible efecto de interacción descrito anteriormente es grave en algunas investigaciones, se duda que sea muy grave en la investigación pedagógica, a condición de que se tomen las precauciones adecuadas. La aplicación de pruebas es una parte normal y aceptada en casi todas las situaciones escolares y universitarias, y por lo mismo no debería tener un profundo efecto sensibilizador. Pese a ello hay ocasiones en que lo tiene. La regla que nos dan Campbell y Stanley es buena: si se van a utilizar procedimientos de pruebas sumamente raros, lo mejor es valerse de diseños que no incluyan prepruebas.

Puntuaciones de diferencia. Vuelva a observar el diseño 19-3, en particular los cambios entre Y_1 y Y_2 . Uno de los problemas más arduos que han atormentado e intrigado a investigadores, especialistas en

medición y estadísticos es la manera de estudiar y analizar dichas puntuaciones. En un libro como el presente es imposible tratar el problema minuciosamente. No obstante, podemos bosquejar las normas y consideraciones de índole general. Se pensaría que es eficaz la aplicación del análisis de varianza a las puntuaciones de diferencia arrojadas por el diseño 19-3 y por otros similares. El análisis puede hacerse si los efectos experimentales son substanciales; pero esas puntuaciones, según vimos en el capítulo 15, normalmente son menos fidedignas que aquéllas a partir de las cuales se calculan. Las diferencias del grupo de control y del grupo experimental pueden carecer de importancia cuando de hecho existe un efecto substancial. Cronbach y Furby incluso llegan a decir que las puntuaciones de cambio no deben utilizarse, que la cuestión esencial es saber si difieren entre sí ambos grupos. (10) ¿Qué hemos, pues, de hacer?

El método generalmente recomendado consiste en utilizar las llamadas puntuaciones regresivas o residuales del incremento. Son las que se calculan por la predicción de las puntuaciones de la posprueba a partir de las puntuaciones de la preprueba en base a la correlación entre ambas y después por la resta de las puntuaciones de la posprueba para obtener las puntuaciones residuales de incremento. (No se preocupe el lector si no comprende bien el método en este momento. Más adelante, luego de estudiar la regresión y el análisis de covarianza, lo entenderá perfectamente.) El efecto de las puntuaciones de la preprueba se suprime de las puntuaciones de la posprueba; en otras palabras, se elimina la influencia de la preprueba en la posprueba. A continuación se verifica la significancia de la diferencia que muestran las medias de tales puntuaciones. Para tal efecto se usa tanto el método que hemos descrito como una ecuación de regresión o el análisis de covarianza, que explicamos en el capítulo 21.

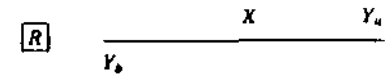
Sin embargo, tampoco es perfecto el empleo de las puntuaciones residuales de incremento ni el análisis de covarianza. Si los sujetos no fueron asignados aleatoriamente al grupo experimental y al grupo de control, el procedimiento no resolverá el problema. Cuando antes del tratamiento experimental los grupos difieren sistemáticamente en otras características relacionadas con la variable dependiente, la manipulación estadística no corregirá semejantes desigualdades. (11) Si se recurre a una preprueba debe utilizarse la aleatorización y el análisis de covarianza, sin olvidar que los resultados han de ser

(10) L. Cronbach y L. Furby, "How Should We Measure 'Change'—or Should We?," *Psychological Bulletin*, LXXIV (1970), 68-80. Se pueden hallar un análisis y un estudio definitivo de las medidas de crecimiento en R. Thorndike, "Intellectual Status and Intellectual Growth," *Journal of Educational Psychology*, LVII (1966), 121-127.

(11) Cronbach y Furby, *op. cit.*, pág. 78.

tratados con especial cuidado. Por último, el análisis de regresión múltiple representa tal vez la mejor solución del problema, según veremos más adelante. (12)

Diseño 19-4: simulación de preprueba y posprueba, con sujetos asignados al azar



El valor de este diseño es dudoso, si bien se le incluye entre los diseños adecuados. Se satisface la exigencia científica de una comparación, pues hay un grupo de comparación (línea inferior). La principal deficiencia del diseño 18-3 (que es una pálida versión de éste) se remedia por medio de la asignación aleatoria. Recuérdese que con el diseño 18-3 fue imposible suponer que los grupos experimental y de control eran equivalentes. En el diseño 19-4 los sujetos han de ser asignados aleatoriamente a ambos grupos, de manera que pueda suponerse que estadísticamente son iguales. Este diseño se utiliza cuando se tienen reservas acerca del efecto reactivo de la preprueba o cuando, por las exigencias prácticas de las situaciones, no queda otra alternativa. Semejante situación se presenta cuando hay oportunidad de ensayar un método o una innovación sólo una vez. Para demostrar la eficacia del método, se suministra una línea base para juzgar el efecto de X en Y y para ello se aplica una preprueba a un grupo similar al grupo experimental. En seguida Y_1 se compara con Y_2 .

La validez de este diseño cae por tierra si uno y otro grupo no se toman aleatoriamente de la misma población o si los sujetos no son asignados al azar a dichos grupos. Aun así, adolece de las deficiencias mencionadas al hablar de otros diseños similares, esto es, otras variables pueden incluir durante el intervalo entre Y_1 y Y_2 . En otras palabras, es supe-

(12) Es lamentable que las complejidades del diseño y del análisis estadístico desahucen al estudiante, al punto que a veces le hacen renunciar a ellos por considerarlos imposibles. Pero la complejidad es propia de la investigación científica: sólo refleja el carácter extraordinariamente intrincado de la realidad psicológica, social y educacional. Esto es a la vez frustrador y estimulante. Al igual que el matrimonio, la investigación de la conducta es ardua y a menudo termina en el fracaso; mas no es imposible. Más aún, es la única manera de obtener un conocimiento confiable de nuestro mundo conductual. Este libro sostiene que el investigador tiene que conocer y entender hasta donde sea posible lo que está haciendo, ser cuidadoso en el diseño y en el análisis, y luego realizar la investigación sin preocuparse demasiado por el aspecto analítico. Lo que debe importarle es el problema de investigación y no su interés por tal aspecto, lo cual no implica que se deba tener una actitud de indiferencia ante el análisis. Sobre todo la investigación pedagógica ha sufrido bastante a causa de ello. Todo esto no significa sino conocimiento, interés y una dosis adecuada de optimismo y escepticismo.

rior al diseño 18-3 pero no se le debe usar si existe una alternativa mejor.

Diseño 19-5: tres grupos sometidos a preprueba y posprueba

R	Y _b	X	Y _a	(Grupo experimental)
	Y _b	~X	Y _c	(Grupo de control 1)
		X	Y _c	(Grupo de control 2)

Este diseño es superior a los precedentes. Además de las ventajas del diseño 19-3, tiene la propiedad de evitar los posibles efectos interactivos que resultan de la preprueba, lo cual se logra por medio del segundo grupo de control (tercera línea). (Aunque parezca un tanto extraño designar a un grupo de control por una X, el de la tercera hilera lo es realmente.) Gracias a las medidas Y_a de que se dispone en este grupo, es posible regular el efecto de interacción. Supongamos que la media del grupo experimental es mucho mayor que la del primer grupo de control (segunda hilera). Quizá no sepamos con certeza si la diferencia fue realmente causada por X. Pudo haber

sido producida por un aumento de sensibilización en los sujetos después de la preprueba y la interacción de su sensibilización con X. Observamos luego la media de Y_a del segundo grupo de control (tercera hilera). También será mucho mayor que la media del primer grupo de control. De ser así, podemos suponer que la preprueba no ha tenido un efecto sensibilizador demasiado grande o que X es lo bastante fuerte para superar cualquier efecto de la interacción sensibilización-X.

Diseño 19-6: cuatro grupos sometidos a preprueba y posprueba (Solomon)

R	Y _b	X	Y _a	(Grupo experimental)
	Y _b	~X	Y _c	(Grupo de control 1)
		X	Y _c	(Grupo de control 2)
		~X	Y _c	(Grupo de control 3)

Este diseño, propuesto por Solomon, (13) es fuerte y satisfactorio desde el punto de vista estético. Está provisto de potentes controles. En realidad, si cambiamos la designación del grupo de control 2 y lo llamamos grupo experimental 2 surgirá una combinación de los diseños 19-3 y 19-1, los dos mejores, en la que el primero forma las dos primeras hileras y el segundo las dos últimas. Las ventajas de uno y otro se combinan en un diseño. Campbell dice que este diseño se ha convertido en el nuevo ideal del científico social. (14) Si bien es una afirmación aventurada, quizás demasiado, indica el gran prestigio de que goza.

Entre las razones por las cuales se afirma que es un diseño firme, se encuentran las siguientes: la

exigencia de comparación queda debidamente satisfecha en las dos primeras hileras y las dos últimas; la asignación aleatoria garantiza la equivalencia estadística de los grupos; la historia y la maduración son controladas por medio de las dos primeras líneas. El posible efecto de interacción causado por la probable sensibilización de los sujetos durante la preprueba, es controlado por las tres primeras hileras. Con la adición de la cuarta línea se controla cualquier efecto contemporáneo transitorio que haya ocurrido entre Y_b y Y_c. Gracias a la combinación de los diseños 19-1 y 19-3, disponemos de la potencia de cada prueba por separado y de la réplica porque en realidad se llevan a cabo dos experimentos. Si la Y_a del grupo experimental fuese mucho mayor que la del grupo de control 1, la del grupo de control 2 sería mucho mayor que la del grupo 3 y al mismo tiempo habría uniformidad de resultados entre ambos experimentos, la evidencia a favor de la validez de nuestra hipótesis de investigación sería contundente.

¿Qué hay de malo en este parangón de diseños? En teoría se ve bien. Parece haber solamente dos dificultades que lo debilitan: una es la practicabilidad: es

	X	~X
Se realizaron prepruebas	Y _a Grupo experimental	Y _c Grupo de control 1
No se realizaron	Y _b Grupo de control 2	Y _d Grupo de control 3

Figura 19-5.

más difícil efectuar dos experimentos simultáneos que uno, además de que el investigador se topa con el problema de encontrar varios sujetos de la misma clase.

La segunda dificultad es el orden estadístico. Obsérvese que no hay equilibrio de grupos. Se dispone de cuatro grupos, mas no de cuatro conjuntos completos de medidas. Las dos primeras hileras, o sea el diseño 19-3, permiten restar Y_b a Y_a y hacer un análisis de covarianza. Con las dos segundas hileras podemos comparar a las Y_c entre sí mediante una prueba t o F, pero el problema reside en obtener un método estadístico global. Una solución sería comparar las Y_c de los grupos de control 2 y 3 con el promedio de las dos Y_b (las dos primeras líneas), y al mismo tiempo comprobar la importancia de la diferencia de las Y_c de las dos primeras hileras. Además, Solomon propuso inicialmente un análisis factorial de varianza del tipo 2 x 2, en que usó los cuatro conjuntos Y_c de medidas. (15) Su idea se bosqueja en la figura 19-5. Un estudio minucioso revelará que constituye un ejemplo excelente del pensamiento investigador, una magnífica combinación de diseño y análisis. Este análisis nos permite estudiar los efectos principales X y ~X, así como a las calificaciones de los sujetos sometidos a preprueba y de los que no lo fueron. Y lo que es más interesante: se puede verificar la interacción de la preprueba y a X, para obtener una respuesta clara al problema anterior.

A pesar que este y otros diseños intrincados ofrecen indudables ventajas, acaso no convenga utilizarlos en forma sistemática. En efecto, sería mejor reservarlos para experimentos de gran trascendencia en los cuales hipótesis ya demostradas con diseños más sencillos vuelven a ser verificadas con mayor rigor y control. En consecuencia, se recomienda que los diseños como el 19-5 y el 19-6 y ciertas variantes de este último, que se verán más adelante, se empleen exclusivamente en pruebas decisivas de la hipótesis de investigación después de realizar algunas experimentaciones preliminares.

Variantes de diseños básicos

Los diseños del 19-1 al 19-6 son los diseños experimentales básicos. Ya hemos indicado algunas de sus variantes. Pueden añadirse grupos experimenta-

les y grupos de control según se necesiten, pero las ideas esenciales son las mismas. Conviene considerar la posibilidad de agregar más grupos. Dentro de ciertos límites este procedimiento aporta más pruebas a favor de la validez de las hipótesis de estudio, como vimos con claridad en el diseño 19-6. Dijimos que era una combinación de otros dos diseños básicos, que en él se aunaban las ventajas de ambos y además el poder de la repetición, así como más controles. De tales ventajas deriva el principio de que, al examinar un diseño de investigación, consideremos la posibilidad de introducir más grupos experimentales como réplicas o variantes de los grupos experimentales y de control.

Importantes variantes del diseño básico son los diseños temporales. La forma del diseño 19-6 puede alterarse para incluir un periodo:

Y _b	X	Y _a
Y _b	~X	Y _c
	X	Y _d
	~X	Y _e

Las Y_d de la tercera y cuarta líneas son observaciones de la variable dependiente en cualquier fecha ulterior especificada. La alteración modifica, claro está, la finalidad del diseño y puede hacer que se pierdan las ventajas del diseño 19-6. Si tuviésemos tiempo, paciencia y recursos, podríamos conservar todos los beneficios anteriores y aun así ampliar el alcance temporal al introducir dos grupos más en el diseño 19-6.

Diseños preexperimentales

Es posible, y a veces necesario, recurrir a diseños que no cumplen con todos los requisitos de la verdadera experimentación. Recuérdese que para realizarla se requieren como mínimo dos grupos: uno que reciba el tratamiento experimental y otro que no lo reciba o bien que lo reciba en forma diferente. Requiere además la manipulación de una variable independiente por lo menos, la asignación aleatoria de los sujetos a los grupos y la asignación al azar de los tratamientos a los grupos. Cuando por una u otra razón se viola uno o varios de los requisitos mencio-

(15) Solomon, op. cit., pág. 146; Campbell, op. cit., pág. 303.

(13) Solomon, op. cit., págs. 137-150. Aunque este diseño puede tener una forma de apareamiento, no la recomendamos. El simbolismo que hemos usado no es el de Solomon.

(14) D. Campbell, "Factors Relevant to the Validity of Experiments in Social Settings," *Psychological Bulletin*, LVI (1957), 297-312. Véase también a D. Campbell y J. Stanley, *Experimental Designs and Quasi-Experimental Designs for Research*, Skokie, Ill.: Rand McNally, 1963.

nados, nos hallamos ante un *diseño preexperimental*. Aunque existen numerosas posibilidades, nos ocuparemos solamente de una en el siguiente apartado.

Diseño preexperimental con grupo experimental y grupo de control. Quizá sea el más usado, pero no nos brinda la seguridad de que ambos grupos sean equivalentes. Ya hemos tratado de su estructura en el diseño 19-3. La forma preexperimental es la siguiente:

Diseño 19-7: forma preexperimental con grupo experimental y grupo de control

Y_1	X	Y_2	(Grupo experimental)
Y_1	$\sim X$	Y_2	(Grupo de control)

Es profunda la diferencia entre los diseños 19-3 y 19-7. En este último los sujetos no se asignan aleatoriamente a los grupos, como en el 19-3 (a), no hay apareamiento de los sujetos, ni asignación al azar como sucede en el 19-3 (b). Por ello, adolece de las deficiencias debidas a la posible falta de equivalencia entre los grupos en variables diferentes de X . Los investigadores procuran establecer la equivalencia por otros medios y en la medida en que lo consiguen el diseño es válido. A continuación se describe cómo lo hacen.

Hay que admitir que a menudo en la investigación es en extremo difícil y a veces imposible igualar los grupos mediante la selección o asignación aleatoria, o mediante el apareamiento. ¿Hemos de renunciar a la investigación? En absoluto. Hay que procurar, en primer lugar, seleccionar y hacer la asignación aleatoria. En caso de que no puedan efectuarse ambos procedimientos, acaso se logre uno de ellos. En caso contrario por lo menos se procurará utilizar muestras de la misma población o lo más semejantes posible. Los tratamientos experimentales han de asignarse aleatoriamente. A continuación se debe probar la semejanza de los grupos por medio de la información disponible: sexo, edad, clase social, etc. La equivalencia de los grupos se comprueba usando las medias y las desviaciones estándar de las prepruebas; serán suficientes las pruebas t y F . También debe determinarse la forma de las distribuciones. Aunque es imposible tener la seguridad que brinda la "aleatorización", si todo lo anterior ha sido demostrado se podrá proseguir el estudio, sabedores de que al menos no hay datos contrarios a la suposición de equivalencia.

Las precauciones referidas acrecientan las posibilidades de alcanzar validez interna. Persisten todavía dificultades, subordinadas todas ellas a una fundamental: la *selección*. (No las comentaremos ahora. En el capítulo previamente citado de la obra

de Campbell y Stanley se da una exposición detallada. Otro valioso libro de consulta es el de Underwood, *Psychological Research*.) (16)

La *selección* es uno de los problemas más arduos de la investigación. Como sus aspectos se exponen detalladamente en el capítulo 22 dedicado a la investigación *ex post facto*, no ofrecemos sino una breve descripción. Uno de los motivos más importantes por los que hacemos hincapié en la selección aleatoria, y en especial en la asignación aleatoria es que deseamos sortear los problemas de selección. Cuando los sujetos son seleccionados y distribuidos en grupos con bases ajenas a los objetivos de la investigación, se dice que se efectúa "selección" o, más exactamente, "autoselección". He aquí un ejemplo: varios voluntarios intervienen en el grupo experimental y otros sujetos participan como miembros del grupo de control. Si los voluntarios difieren en una característica relacionada con Y , la variable dependiente, es posible que la diferencia final entre ambos grupos de deba a esta característica y no a X . Si el experimento tiene cierta clase de aprendizaje como variable dependiente, sin duda los primeros lograrán un mejor rendimiento en Y , en virtud de su inteligencia superior y pese a la semejanza inicial entre ambos grupos durante la preprueba. (Adviértase que si nos hubiésemos valido exclusivamente de voluntarios, y los hubiésemos asignado aleatoriamente a los dos grupos, desaparecería la dificultad de selección pero decrecería la validez externa o representatividad.)

Otro ejemplo más frecuente en la investigación pedagógica es distribuir algunos grupos escolares entre el grupo experimental y el de control. Si un número bastante grande de ellos se selecciona y se asigna aleatoriamente a uno y otro grupo, no habrá gran problema. Pero si no se asignan de esta manera algunos escogerán los grupos experimentales y quizá posean características que los predispongan a obtener puntuaciones promedio Y más altas que las de los demás. Por ejemplo, sus profesores pueden ser más listos, inteligentes y agresivos. Las características interactúan con la selección para producir, cualquiera que sea la X , puntuaciones de Y mayores en el grupo experimental que en el grupo de control. En otras palabras, lo que influye en el proceso selectivo, como sucede con los sujetos voluntarios, influye del mismo modo en las medidas de la variable dependiente. Esto acontece aunque la preprueba demuestre que los grupos son idénticos en la variable dependiente. Esta manipulación de X es "eficaz" pero su eficacia no es absoluta, sino que depende de la selección o *autoselección*.

Diseños temporales

Un problema común de investigación, especialmente en los estudios dedicados al desarrollo y cre-

(16) Underwood, *op. cit.*, caps. 4 y 5.

cimiento del niño, lo plantea el estudio de individuos y grupos en los que se usa el tiempo como variable. Son los estudios longitudinales de sujetos, generalmente de niños, en distintas etapas de su vida. He aquí uno de dichos diseños:

Diseño 19-8: diseño temporal longitudinal

Y_1 Y_2 X Y_3 Y_4

Nótese su semejanza con el diseño 18-2, en que al grupo se le compara consigo mismo. El diseño 19-8 nos permite sortear una de las dificultades del diseño 18-2. Su empleo nos permite distinguir los aspectos reactivos de la medición y el efecto de X . Además, gracias a él sabemos si, en caso que las mediciones produzcan un efecto reactivo, X tiene un efecto que supere a éste con creces. El efecto reactivo se manifestará en Y_3 ; y ésta puede contrastarse con Y_2 . Si se da un incremento en Y_3 que rebase totalmente al incremento en Y_2 , debe atribuirse a X . Un argumento parecido se aplica a la maduración y a la historia.

Un problema de los estudios longitudinales o temporales, especialmente tratándose de niños, es el crecimiento y el aprendizaje que se realizan con el tiempo. Los niños no dejan de crecer y aprender para facilitar la investigación. Cuanto más largo el periodo, mayor será el problema. En otras palabras, el tiempo es ya una variable en cierto modo. En un diseño como el 18-2, Y_1 , X , Y_2 , la variable "tiempo" puede hacer confusa a X (la variable independiente experimental). Si existe una apreciable diferencia entre Y_1 y Y_2 será imposible saber si X o una variable temporal originó el cambio. Al contrario, con el diseño 19-8 se dispone de otras medidas de Y y, por lo tanto, de una línea base con la cual comparar el cambio operado en Y que se atribuye a X .

Campbell y Stanley piensan que la historia es el problema más difícil de este diseño. La cuestión es que no fue X la que produjo un cambio en Y , sino algún otro hecho o combinación de hechos que acontecieron durante el periodo experimental. La historia es un problema, pero si otras X están operando deberían manifestarse en el intervalo, digamos entre Y_1 y Y_2 , lo mismo que entre Y_3 y Y_4 . Si además de la X experimental ocurren sin cesar hechos o variables extrañas, es evidente que la historia mermará la validez interna de un estudio temporal. Los investigadores deben estar muy atentos ante otras posibilidades del mismo tipo a fin de impedir su aparición, y deben tomarlas en cuenta en la interpretación de resultados o demostrar que la manipulación experimental X supera esas influencias extrañas.

Tomemos un ejemplo pedagógico para ilustrar la investigación temporal y las posibles influencias extrañas. Supongamos que un consejo técnico, pro-

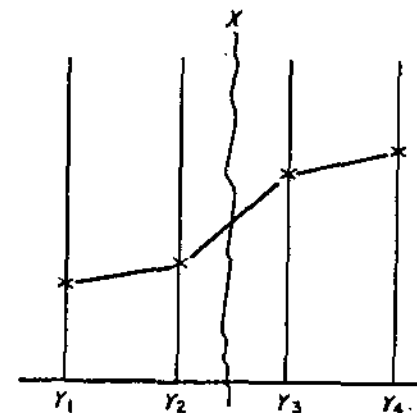


Figura 19-6.

cupado por la moral de su profesorado, instituye un nuevo sistema tendiente a mejorarla. El personal administrativo dispone de dos medidas de la moral del profesorado tomadas a intervalos de tres meses: Y_1 y Y_2 . Se implanta el nuevo sistema, X . La moral del profesorado se vuelve a medir en dos ocasiones posteriores: Y_3 y Y_4 . Supongamos que la gráfica de los resultados es la que aparece en la figura 19-6 (conviene graficar los datos temporales). Se aprecia que el efecto de X parece ser muy superior al efecto del tiempo. Si bien la moral parecía haber aumentado antes de introducir X , ocurrió un ascenso relativamente pronunciado después de introducirla. Luego prosiguió el ritmo anterior de incremento. El único problema es: ¿ocurrió otra cosa entre Y_2 y Y_3 que causó el incremento de moral? Es poco probable, pero hemos de examinar cuidadosamente la situación. Tal vez el inspector tomó alguna medida al mismo tiempo que el consejo técnico cambió el sistema en cuestión.

Es fácil transferir los ejemplos citados a los estudios sobre los datos del crecimiento y aprendizaje en el niño. Las Y serían las mediciones de habilidad para la lectura; X sería un nuevo método para enseñar a leer. Resultados como los de la figura 19-6 le parecerían alentadores al maestro. Hemos de ser aún más cautos. Si se realiza un estudio sobre el aprendizaje, son indispensables uno o más grupos de control, por las razones ya consideradas y también por el efecto de Hawthorne. (17) Prácticamente

(17) El nombre del efecto proviene de la planta Hawthorne de la Western Electric Corporation, donde se observó por primera vez. Una descripción breve y de fácil consulta acerca de una parte de dicha investigación se encuentra en G. Homans, "Group Factors in Worker Productivity," en H. Proshansky y B. Seidenberg, eds., *Basic Studies in Social Psychology*, Nueva York: Holt, Rinehart and Winston, 1965, págs. 592-604.

cualquier cambio, cualquier atención adicional, cualquier manipulación experimental (o aun sin ella) y hasta el simple conocimiento de que se está efectuando un estudio bastan para que los sujetos cambien.

En resumen, si prestamos atención a la gente, ésta responderá. En el ejemplo, quizá cualquier método, con tal que fuese distinto de los que se utilizan ordinariamente en el aula, haría que aumentasen las puntuaciones de lectura y de ahí la necesidad de contar con un grupo de control al cual se le aplique otro método de placebo. (Resulta difícil pensar en un grupo de control en la situación técnico-moral del profesorado.)

El análisis estadístico de las medidas temporales plantea un problema especial y arduo: las pruebas normales de significancia pueden arrojar resultados espurios cuando se utilizan con medidas temporales. Una razón es que esos datos tienden a ser sumamente variables y es fácil interpretar los cambios no debidos a *X* como causados por ella. Es decir, con el tiempo los datos, las puntuaciones individuales y sus promedios tienden a variar considerablemente. Con frecuencia se cae en la trampa de conceder importancia a uno de esos cambios especialmente si coincide con nuestra hipótesis. Si podemos suponer legítimamente que otras influencias además de la *X*, tanto sistemáticas como aleatorias, son uniformes en toda la serie de las *Y*, será posible resolver el problema estadístico. (18) Pero semejante suposición puede ser, y probablemente muchas veces lo es, injustificada.

El investigador que realiza estudios de tiempo debe hacer un estudio especial sobre los problemas estadísticos y consultar a un experto. Para el experimentador esta complejidad estadística es negativa, en el sentido de que acaso lo aleja de los estudios prácticos que se requieren. Los diseños longitudinales de grupos aislados son particularmente aptos para la investigación de clases individuales; por ello se recomienda que en los estudios longitudinales de métodos o en los estudios de niños colocados en situaciones pedagógicas, el análisis se limite a trazar gráficas de los resultados e interpretarlas cualitativamente. Sin embargo, las pruebas decisivas, en especial las destinadas a la publicación, han de estar respaldadas con pruebas estadísticas.

Desde luego, el diseño 19-8 presenta algunas variaciones. Una de importancia consiste en añadir uno o varios grupos de control; otra en introducir más observaciones temporales; una tercera en agregar más *X*, esto es, más intervenciones experimentales.

(18) Una explicación estadística se encuentra en A. Edwards, *Experimental Design in Psychological Research*, 3a. ed. Nueva York: Holt, Rinehart and Winston, 1968, cap. 14.

Observaciones finales

Es importante recordar que los diseños de este capítulo son generales: han sido reducidos a sus elementos esenciales a fin de descubrir sus estructuras. Teniendo muy en cuenta esas estructuras, el estudiante estará en condiciones de usar los diseños más específicos del análisis de varianza y paradigmas afines. Es verdad que se pueden estudiar sólo los diseños más específicos, pero el conocimiento de los diseños generales aumenta la flexibilidad en el sentido de que es más fácil adaptarlos a las necesidades de la investigación. Este punto se comentó en el capítulo 13, donde se habló de las comparaciones planeadas y de las pruebas post hoc. En todo caso, las relaciones entre los paradigmas generales y específicos de los diseños se examinarán en los capítulos 20 y 21.

Un segundo punto es que las ideas y paradigmas del diseño básico nacieron de consideraciones experimentales. También en este caso el ideal de la investigación científica lo constituye el experimento controlado. Pero como hemos señalado una y otra vez, gran parte de la investigación de la conducta es de carácter no experimental y hay que analizar sus datos. Cuanto más nos aproximemos al diseño experimental controlado, tanto mejor. En consecuencia, el diseño creado para la investigación experimental es el ideal, la norma. En la medida en que sea preciso renunciar a los criterios del diseño experimental, se debilitarán los resultados y conclusiones de la investigación. Y de ahí que constantemente volvamos la vista hacia el ideal aunque nos sepamos incapaces de alcanzarlo. Lo importante es que no renunciemos fácilmente a nuestros principios y criterios, trátase de una investigación experimental o no experimental.

Tercero: lo importante en el estudio y en el empleo de los diseños es conocer sus principios, objetivos, ventajas y deficiencias. El conocimiento de los detalles de muchos diseños y sus posibilidades no es, ni con mucho, tan importante como conocer su naturaleza general y su aplicabilidad. Los problemas y los estudios difieren mucho y al investigador le es necesario adaptar los principios y paradigmas generales a las exigencias de los problemas específicos. Por ejemplo, un diseño factorial de tipo 2×2 puede concebirse como un análisis de varianza de una sola variable con cuatro grupos, con comparaciones planeadas. Como siempre, la manera de concebirlo y manejarlo depende de las preguntas planteadas.

Finalmente, es necesario saber que los investigadores de la conducta muestran dos fuertes tendencias: una hacia la investigación experimental y otra hacia la investigación no experimental. Hay quien dice que la investigación del comportamiento debe ser experimental y quien sostiene que los experimentos son absurdos. Probablemente la mitad de la

investigación psicológica y pedagógica, así como la mayor parte de la investigación sociológica y antropológica no sea experimental. Algunos investigadores de psicología, sociología y educación llegan incluso a afirmar que los problemas más trascendentes e interesantes no se prestan a un método experimental.

Lo ideal sería que siempre que fuese posible abordásemos los problemas y verificásemos las hipótesis con métodos experimentales y no experimentales. Si una variable independiente es manipulable, puede y debe usarse el método experimental. Sin embargo, algunas variables importantes no se pueden estudiar experimentalmente porque no son manipulables, al menos en nuestra sociedad: inteligencia, educación del niño, valores religiosos, honradez, características de los maestros, ambiente del hogar y muchas otras. Quizá algunas puedan manipularse, pero su misma naturaleza lo impide en la mayoría de los casos. Muchas otras son intrínsecamente manipulables: refuerzo, ejercicios, métodos de enseñanza, ambientes de escuela y de clase, y ciertas conductas de los maestros. Las hay que son mensurables y manipulables a la vez, tales como la ansiedad y la frustración. En conclusión, su misma multiplicidad y complejidad nos indica que es erróneo insistir en la preponderancia de la investigación experimental o no experimental en la educación, psicología y sociología.

Guías de estudio

1. La frase con que se inició este capítulo es: "El diseño es la norma que regula los datos." ¿Qué significa? Justifíquela.
2. Imagine que usted es un psicopedagogo y planea comprobar la hipótesis de que si se retroalimenta a los maestros con información psicológica aumentará su entendimiento acerca de los niños y en consecuencia se incrementará el aprendizaje de éstos. Bosqueje un diseño *ideal*

con el cual comprobar la hipótesis, suponiendo que tiene control absoluto sobre la situación y suficiente dinero y ayuda. (*Son condiciones importantes que se incluyen para liberar al lector de las limitaciones prácticas que con tanta frecuencia vician los buenos diseños.*) Elabore dos diseños: uno con asignación aleatoria completa, ambos según el paradigma del diseño 19-1. En uno use únicamente una variable independiente y el análisis unidireccional de varianza. En el otro, utilice dos variables independientes y un diseño factorial sencillo.

¿Cuál de los dos es más satisfactorio en cuanto a sus poderes reguladores y a la información que su ministra? ¿Cuál demuestra mejor la hipótesis? ¿Por qué?

3. Elabore una investigación para demostrar la hipótesis de la sugerencia anterior, pero esta vez modifique el diseño omitiendo la asignación aleatoria. Compare la eficacia relativa de ambos. ¿Cuál de los dos merecerá más confianza? ¿Por qué? Explíquelo en detalle.
4. En un estudio sobre la relación de las actitudes de niños hacia la escuela y hacia la orientación vocacional efectuado en una secundaria, el investigador escogió 30 alumnos del octavo grado, quienes debido a problemas disciplinarios habían sido enviados a recibir orientación vocacional en el año anterior; luego igualó a cada uno en sexo e inteligencia con otro alumno del octavo grado que no hubiese sido enviado a recibirla. Al inicio y al final del año comparó las actitudes de ambos grupos frente a la escuela y encontró una notable diferencia al inicio del año, no así al final. Llegó a la conclusión de que la orientación vocacional tiene un efecto saludable en las actitudes de los niños hacia la escuela.

Haga una crítica de la investigación: ¿es buena? ¿Es mala? ¿Muy buena? ¿Muy mala? Tenga presente los siguientes elementos: muestreo, comparabilidad de grupos y control.