



CAPÍTULO 18

DISEÑO DE INVESTIGACIÓN: PROPÓSITO Y PRINCIPIO

- **PROPÓSITOS DEL DISEÑO DE INVESTIGACIÓN**
 - Un ejemplo
 - Un diseño más fuerte
- **EL DISEÑO DE INVESTIGACIÓN COMO CONTROL DE LA VARIANZA**
 - Un ejemplo controversial
- **MAXIMIZACIÓN DE LA VARIANZA EXPERIMENTAL**
- **CONTROL DE VARIABLES EXTRAÑAS**
- **MINIMIZACIÓN DE LA VARIANZA DEL ERROR**

El *diseño de investigación* constituye el plan y la estructura de la investigación, y se concibe de determinada manera para obtener respuestas a las preguntas de investigación. El plan es el esquema o programa general de la investigación; incluye un bosquejo de lo que el investigador hará, desde formular las hipótesis y sus implicaciones operacionales hasta el análisis final de los datos. La estructura de la investigación resulta más difícil de explicar, ya que el término *estructura* presenta dificultad para ser definido claramente y sin ambigüedades. A causa de que es un concepto que irá tomando gran importancia conforme se continúe el estudio, se realizará una pausa para intentar definirlo y ofrecer una breve explicación. En este momento la disertación será necesariamente un poco abstracta, sin embargo, ejemplos posteriores serán más concretos. Más importante aún, el concepto se encontrará poderoso, útil e incluso indispensable, especialmente en el estudio posterior del análisis multivariado, donde el concepto "estructura" es clave, y cuyo entendimiento se vuelve esencial para comprender la mayoría de la metodología de investigación contemporánea.

Una *estructura* es el marco de referencia, la organización o configuración de los elementos de la estructura, relacionados en formas específicas. La mejor forma de especificar una estructura consiste en escribir una ecuación matemática que relacione las partes de la estructura entre sí. Dicha ecuación matemática, puesto que sus términos están definidos y relacionados específicamente por la ecuación (o conjunto de ecuaciones), no es ambigua.

En resumen, una estructura es un paradigma o modelo de las relaciones entre las variables de un estudio. Los términos *estructura*, *modelo* y *paradigma* son problemáticos debido a que es difícil definirlos con claridad y sin ambigüedades. Un "paradigma" es un modelo, un ejemplo. Los diagramas, gráficas y bosquejos verbales son paradigmas. Aquí se utiliza "paradigma" en lugar de "modelo" porque "modelo" tiene otro importante significado en la ciencia, significado al que se regresará en el capítulo 37, cuando se analice la comprobación de una teoría utilizando procedimientos multivariados y "modelos" de aspectos de teorías.

Un diseño de investigación expresa tanto la estructura del problema de investigación como el plan de investigación utilizado para obtener evidencia empírica sobre las relaciones del problema. Pronto se presentarán ejemplos del diseño y de la estructura que quizás animen esta discusión abstracta.

Propósitos del diseño de investigación

El diseño de investigación incluye dos propósitos básicos: 1) *proporcionar respuestas a preguntas de investigación* y 2) *controlar la varianza*. El diseño ayuda a los investigadores a obtener respuestas a las preguntas de investigación, y también a controlar las varianzas experimental, extraña y del error del problema de investigación particular en estudio. Ya que puede decirse que toda actividad de investigación tiene el propósito de generar respuestas a preguntas de investigación, es posible omitir este propósito en el análisis y de afirmar que el diseño de investigación tiene un propósito fundamental: controlar la varianza. Sin embargo, tal delimitación del propósito del diseño es peligrosa. Sin un fuerte énfasis en las preguntas de investigación y en el uso del diseño para ayudar a proporcionar respuestas a dichas preguntas, el estudio del diseño puede degenerar en un ejercicio técnico interesante, pero estéril.

Los diseños de investigación se inventaron para permitir a los investigadores responder preguntas de la forma más válida, objetiva, precisa y económica posible. Los planes de investigación se conciben de forma deliberada y específica, y son ejecutados para obtener evidencia empírica que apoye al problema de investigación. Los problemas de investigación pueden ser, y son, expresados en forma de hipótesis; éstas se formulan en un momento de la investigación de manera que puedan ser probadas empíricamente. Los diseños se elaboran con cuidado para que proporcionen respuestas confiables y válidas a las preguntas de investigación contenidas en las hipótesis. Es posible realizar una sola observación e inferir que la relación hipotetizada existe, con base en esta única observación; pero es evidente que no se puede aceptar la inferencia realizada de esa forma. Por otro lado, también es factible realizar cientos de observaciones e inferir que la relación hipotetizada existe, con base en estas múltiples observaciones, en cuyo caso se puede o no aceptar como válida la inferencia. El resultado depende de la manera en que se hicieron las observaciones y la inferencia. Un diseño planeado y ejecutado de forma adecuada ayuda en mucho a permitirse confiar tanto en las observaciones como en las inferencias.

¿Cómo logra esto el diseño? El diseño de investigación establece el marco de referencia para el estudio de las relaciones entre variables. Indica, en cierto sentido, qué observaciones hacer, cómo hacerlas y cómo realizar las representaciones cuantitativas de las observaciones. Estrictamente hablando, el diseño no "dice" precisamente qué hacer, sino qué "sugiere" la dirección de cómo realizar las observaciones y el análisis. Un diseño adecuado "sugiere", por ejemplo, cuántas observaciones deben efectuarse y qué variables son activas y cuáles son atributivas. Entonces se actúa para manipular las variables activas y categorizar y medir las variables atributivas. Un diseño indica qué tipo de análisis estadís-

tico emplear. Por último, un diseño adecuado bosqueja las conclusiones que posiblemente se obtengan del análisis estadístico.

Un ejemplo

Se ha dicho que los colegios y las universidades discriminan a las mujeres respecto a los procesos de contratación y de admisión. Suponga que se desea probar la discriminación en la admisión. La idea para este ejemplo proviene del inusual y extraño experimento de Walster, Cleary y Clifford (1970) citado anteriormente. Se diseña un experimento de la siguiente manera: se envían solicitudes de admisión a una muestra aleatoria de 200 colegios, basando las solicitudes en varios casos modelo seleccionados sobre un rango de habilidades probadas, con todos los detalles iguales excepto el género. La mitad de las solicitudes serán de hombres, y la otra mitad, de mujeres. Manteniendo las otras cuestiones iguales, se espera aproximadamente igual número de aceptaciones y de rechazos; entonces la *aceptación* es la variable dependiente, la cual se mide con una escala de tres puntos: aceptación completa, aceptación con reservas y rechazo. Llámese a los hombres A_1 y a las mujeres A_2 . El paradigma del diseño se presenta en la figura 18.1.

El diseño es el más simple posible, dados los requerimientos mínimos de control. Los dos tratamientos se asignan a los colegios aleatoriamente. Entonces, cada colegio recibirá una solicitud, ya sea de un hombre o de una mujer. Se probará la significancia estadística de la diferencia entre las medias M_{A_1} y M_{A_2} con un prueba t o F . La hipótesis sustantiva es: $M_{A_1} > M_{A_2}$, o se admitirán más hombres que mujeres. Si no hay discriminación en la admisión, entonces M_{A_1} sería estadísticamente igual a M_{A_2} . Suponga que una prueba F indica que las medias no son significativamente diferentes. ¿Se puede estar seguro de que no hay práctica de discriminación (en promedio)? Mientras que el diseño de la figura 18.1 es satisfactorio hasta ahora, quizá no llega suficientemente lejos.

Un diseño más fuerte

Walster y sus colegas utilizaron otras dos variables independientes, la *raza* y la *habilidad*, en un diseño factorial. En el ejemplo se eliminó *raza* —no fue estadísticamente significativa ni interactuó de manera significativa con otras variables— y se enfatizó el género y la habilidad. Si un colegio basa su selección de estudiantes de nuevo ingreso estrictamente en las habilidades, entonces no hay discriminación (a menos, por supuesto, que la selección por habilidades se considere discriminación). Añádase *habilidad* al diseño de la figura 18.1 usando tres niveles; es decir, además de designar a las aplicaciones como *hombre* y *mujer*, también se designan como *habilidad alta*, *habilidad media* y *habilidad baja*. Por ejemplo, tres de los solicitantes pueden ser: hombre con habilidad media, mujer con habilidad alta y

FIGURA 18.1

Tratamientos	
A_1 (Hombre)	A_2 (Mujer)
Puntuaciones de aceptación	
M_{A_1}	M_{A_2}

 FIGURA 18.2

		Género		
		A_1 (Hombre)	A_2 (Mujer)	
Habilidad	B_1 (Alta)			M_{B_1}
	B_2 (Media)	Puntuaciones de aceptación		M_{B_2}
	B_3 (Baja)			M_{B_3}
		M_{A_1}	M_{A_2}	

mujer con habilidad baja. Ahora, si no existen diferencias significativas entre los géneros ni la interacción de *género* y *habilidad* es significativa, ésta sería una evidencia considerablemente más fuerte de que no hay discriminación que la proporcionada por el diseño y por la prueba estadística de la figura 18.1. Ahora se utiliza el diseño ampliado para explicar esta afirmación y analizar ciertos aspectos del diseño de investigación. El diseño ampliado se presenta en la figura 18.2.

El diseño es un factorial de 2×3 . Una variable independiente, A , es el género, la misma que en la figura 18.1. La segunda variable independiente, B , es la habilidad, que se manipuló para indicar, de varias formas, cuáles son los niveles de habilidad de los estudiantes. Es importante no confundirse por el nombre de las variables; *género* y *habilidad* son por lo común variables atributivas, por lo tanto, no experimentales. Sin embargo, en este caso se manipulan. Los registros de los estudiantes enviados a los colegios fueron sistemáticamente ajustados para que se adecuaran a las seis casillas de la figura 18.2. Por ejemplo, un caso en la casilla A_1B_2 , sería el registro de un hombre con habilidad media, que es el registro que el colegio evalúa para la admisión.

Suponga que se piensa que la discriminación en contra de las mujeres toma una forma más sutil que la simple exclusión a todos los niveles: se piensa que se discrimina contra las mujeres con habilidad baja (en comparación con los hombres). Ésta es una hipótesis de interacción. De cualquier manera, se utiliza este problema y el paradigma de la figura 18.2 como base para analizar algunos elementos del diseño de investigación.

Los problemas de investigación sugieren diseños de investigación. Puesto que la hipótesis antes discutida es de interacción, evidentemente un diseño factorial es el apropiado. A es el *género*; B es la *habilidad*; A se divide en A_1 y A_2 y B en B_1 , B_2 y B_3 .

El paradigma de la figura 18.2 sugiere varias cosas. La primera, y la más obvia, es que se requiere un gran número de participantes; específicamente se necesitan $6n$ participantes (n es igual al número de sujetos en cada casilla). Si se decide que n debe ser 20, entonces se requiere tener 120 sujetos para el experimento. Observe aquí la "sabiduría" del diseño; si tan sólo se estuvieran probando los tratamientos y se ignorara la habilidad, únicamente se necesitarían $2n$ sujetos. Es preciso observar que algunos autores como Simon (1976, 1987); Simon y Roscoe (1984) y Daniel (1976) discrepan con este enfoque para todo tipo de problemas. Ellos consideran que muchos diseños contienen réplicas ocultas y que serían suficientes mucho menos de 20 participantes por casilla. Tales diseños requieren una planeación mucho más cuidadosa; pero el investigador puede obtener información mucho más útil y estudiar más variables independientes en lugar de sólo dos o tres.

Existen formas para determinar el número de participantes que se requieren en un estudio. Tal determinación forma parte del "poder", que se refiere a la habilidad de una prueba de significancia estadística para detectar diferencias en las medias (u otros estadísticos), cuando en realidad existen tales diferencias. En el capítulo 8 se explica el tamaño de

las muestras y su relación con la investigación. Sin embargo, el capítulo 12 presenta un método para estimar el tamaño de las muestras de manera que se cumplan ciertos criterios. El poder es un valor fraccional entre 0 y 1.00 que se define como $1 - \beta$, donde β es la probabilidad de cometer un error tipo II, el cual sucede cuando no se rechaza una hipótesis nula falsa. Si el poder es alto (cercano a 1.00) indica que si la prueba estadística no fue significativa, la investigación sugiere que la hipótesis nula es verdadera. El poder también indica qué tan sensible es la prueba estadística para detectar diferencias reales. Si la prueba estadística no es lo suficientemente sensible para hacer esto, se dice que la prueba tiene poco poder. Una prueba altamente sensible, que puede detectar diferencias verdaderas, se considera de alto poder. En el capítulo 16 se analizó la diferencia entre las pruebas estadísticas paramétricas y las no paramétricas. Las pruebas no paramétricas son generalmente menos sensibles que las pruebas paramétricas; como resultado, se considera que las primeras tienen menos poder que las segundas. Uno de los libros más completos sobre la cuestión de la estimación del poder es el de Cohen (1988). Jaccard y Becker (1997) ofrecen una introducción fácil de entender al análisis del poder.

En segundo lugar, el diseño indica que los “participantes” (en este caso los colegios) pueden asignarse aleatoriamente tanto a A como a B , ya que ambas son variables experimentales. Sin embargo, si *habilidad* fuese una variable no experimental atributiva, entonces los participantes podrían ser asignados de manera aleatoria a A_1 y A_2 , pero no a B_1 , B_2 ni B_3 .

En tercer lugar, de acuerdo al diseño, las observaciones realizadas en los “participantes” deben realizarse de manera independiente. La puntuación de un colegio no debe afectar a la puntuación de otro. Reducir el diseño a un bosquejo como el que se indica en la figura 18.2, en efecto, prescribe las operaciones necesarias para obtener las medidas apropiadas para el análisis estadístico. Una prueba F depende del supuesto de la independencia de las medidas de la variable dependiente. Si aquí *habilidad* es una variable atributiva y a los individuos se les mide la inteligencia, por ejemplo, entonces el requisito de independencia está en mayor riesgo debido a la posibilidad de que un sujeto vea los documentos de otro y a que los maestros “ayuden” inconscientemente o conscientemente a los estudiantes con las respuestas, entre otras razones. Los investigadores tratan de prevenir este tipo de situaciones, no tanto por razones morales sino para satisfacer los requisitos de un diseño y una estadística sólidos.

Un cuarto punto resulta bastante obvio ahora: la figura 18.2 sugiere un análisis factorial de varianza, pruebas F , medidas de asociación y, quizá, pruebas *post hoc*. Si la investigación está bien diseñada antes de la recolección de los datos —como en realidad lo hicieron Walster *et al.*— la mayoría de los problemas estadísticos pueden resolverse. Además, se evitan ciertos problemas molestos antes de que surjan, o incluso pueden prevenirse del todo. Sin embargo, con un diseño inadecuado, los problemas referentes a las pruebas estadísticas apropiadas se vuelven muy molestos. Una de las razones del gran énfasis de este libro en tratar los problemas de diseño y estadísticos de forma concomitante, es que esto permite señalar maneras de evitar tales problemas. Si el diseño y el análisis estadístico se planean simultáneamente, el trabajo analítico se volverá sencillo y ordenado.

Un dividendo bastante útil del diseño es el siguiente: un diseño claro, como el de la figura 18.2, sugiere qué prueba estadística realizar. Por ejemplo, un diseño aleatorio simple de una variable con dos particiones o tratamientos, A_1 y A_2 , permite tan sólo una prueba estadística de la diferencia entre los dos estadísticos producidos por los datos. Dichos estadísticos pueden ser dos medias, dos medianas, dos rangos, dos varianzas, dos porcentajes, etcétera. Sólo una prueba estadística es generalmente posible. Sin embargo, con el diseño de la figura 18.2 existen tres pruebas estadísticas posibles: 1) entre A_1 y A_2 ; 2) entre B_1 , B_2 y B_3 , y 3) la interacción entre A y B . En la mayoría de las investigaciones, no

 FIGURA 18.3

Condición B_1 Tratamientos		Condición B_2 Tratamientos	
A_1	A_2	A_1	A_2
M_{A_1}	M_{A_2}	M_{A_1}	M_{A_2}

todas las pruebas estadísticas tienen la misma importancia; las importantes, en efecto, son aquellas directamente relacionadas con los problemas e hipótesis de investigación.

En el presente caso, la hipótesis de interacción [la del inciso 3) anterior] es la más importante, ya que se supone que la discriminación depende del nivel de *habilidad*. Los colegios quizá discriminen a diferentes niveles de habilidad. Como se sugirió antes, las mujeres (A_2) tal vez sean aceptadas más que los hombres (A_1) en el nivel de habilidad más alto (B_1); mientras que quizá sean menos aceptadas en el nivel de habilidad más bajo (B_2). Debería ser evidente que el diseño de investigación no es estático. El tener conocimiento sobre diseño puede ayudar a planear y realizar mejor investigación, y también puede sugerir la comprobación de hipótesis. Y quizá más importante: puede llevar a que uno se dé cuenta de que el diseño de un estudio no es adecuado a las demandas planteadas. ¿Qué significa esta afirmación un tanto peculiar?

Suponga que se formula la hipótesis de interacción como se bosquejó anteriormente, sin saber nada sobre el diseño factorial; en realidad se establece un diseño que consiste de dos experimentos, en uno de los cuales se prueba A_1 contra A_2 , bajo la condición B_1 . En el segundo experimento se prueba A_1 contra A_2 , bajo la condición B_2 . El paradigma se vería como el que se muestra en la figura 18.3. (Para simplificar las cosas, únicamente se utilizan dos niveles de B : B_1 y B_2 ; por lo tanto, el diseño se reduce a uno de 2×2 .)

El punto importante a señalar es que no es posible realizar una prueba *adecuada* de la hipótesis con este diseño. A_1 puede probarse contra A_2 bajo las dos condiciones B_1 y B_2 para asegurarse. Pero no es posible saber con claridad y sin ambigüedades, si existe una interacción significativa entre A y B . Aun cuando $M_{A_1} > M_{A_2}$ | B_1 (M_{A_1} es mayor que M_{A_2} , bajo la condición B_1), como se hipotetizó, el diseño no puede ofrecer una clara posibilidad de confirmación de la interacción hipotetizada, debido a que no se puede obtener información sobre las diferencias entre A_1 y A_2 en los dos niveles de B (B_1 y B_2). Recuerde que una hipótesis de interacción implica, en este caso, que la diferencia entre A_1 y A_2 es distinta en B_1 de lo que es en B_2 . En otras palabras, la información tanto de A como de B *juntas en un experimento* es necesaria para probar una hipótesis de interacción. Si los resultados estadísticos de experimentos separados mostraran una diferencia significativa entre A_1 y A_2 en un experimento bajo la condición B_1 , y no mostraran diferencias significativas en otro experimento bajo la condición B_2 , entonces hay *presunta* evidencia de que la hipótesis de interacción es correcta. Pero no es suficiente contar con presunta evidencia especialmente cuando se sabe que es posible obtener una mejor evidencia.

Suponga que en la figura 18.3, las medias de las casillas fueran, de izquierda a derecha: 30, 30, 40, 30. Tal resultado parecería apoyar la hipótesis de interacción, ya que hay una diferencia significativa entre A_1 y A_2 en el nivel B_2 , pero no en el nivel B_1 . Pero no puede tenerse la certeza de que esto es así, incluso si la diferencia entre A_1 y A_2 es estadísticamente significativa. La figura 18.4 presenta cómo resultaría esto si se hubiese utilizado un diseño factorial. (Las cifras en las casillas y en los márgenes son medias.) Considerando que los efectos principales, A_1 y A_2 ; B_1 y B_2 , fueran significativos, todavía es posible que la interac-

FIGURA 18.4

	A_1	A_2	
B_1	30	30	30
B_2	40	30	35
	35	30	

ción no sea significativa. A menos que la hipótesis de interacción se pruebe específicamente, la evidencia para determinar la interacción es mera presunción, ya que falta la prueba estadística de la interacción que un diseño factorial proporciona. Debe quedar claro que el conocimiento sobre diseño hubiese mejorado este experimento.

El diseño de investigación como control de la varianza

La principal función técnica del diseño de investigación es *controlar la varianza*. Un diseño de investigación constituye, por así decirlo, un conjunto de instrucciones para que el investigador reúna y analice los datos de cierta forma; por lo tanto, es un mecanismo de control. El principio estadístico que subyace a este mecanismo, como se dijo antes, es: *maximizar la varianza sistemática, controlar la varianza sistemática extraña y minimizar la varianza del error*. En otras palabras, se debe *controlar* la varianza.

De acuerdo con este principio, al construir un diseño de investigación eficiente, el investigador intenta: 1) maximizar la varianza de la variable o variables de la hipótesis sustantiva de investigación, 2) controlar la varianza de variables extrañas o "indeseables" que puedan tener un efecto en los resultados experimentales y 3) minimizar la varianza del error o aleatoria, incluyendo los llamados errores de medición. Ahora se verá un ejemplo.

Un ejemplo controversial

La controversia abunda en toda la ciencia y parece ser especialmente rica y variada en las ciencias del comportamiento. Dos controversias han surgido a partir de diferentes teorías del comportamiento y aprendizaje humanos. Los teóricos del reforzamiento han demostrado ampliamente que el reforzamiento positivo puede incrementar el aprendizaje. Sin embargo, como siempre, las cuestiones no son tan simples. El supuesto efecto benéfico de las recompensas externas se ha cuestionado; la investigación ha mostrado que la recompensa extrínseca puede tener una influencia perjudicial en la motivación, interés intrínseco y aprendizaje de los niños. En los años setenta, se publicó una serie de artículos y estudios que mostraban los posibles efectos dañinos del uso de la recompensa. En uno de dichos estudios, Amabile (1979) demostró que la evaluación externa tiene un efecto perjudicial sobre la creatividad artística. Otros estudios incluyen el de Deci (1971) y el de Lepper y Greene (1978). Al mismo tiempo, incluso el principio del reforzamiento en apariencia simple, no es tan simple. Sin embargo, en años recientes han aparecido varios artículos que defienden los efectos positivos de la recompensa (véase Eisenberger y Cameron, 1996; Sharpley, 1988; McCullers, Fabes y Moran, 1987; Bates, 1979).

Existen diversas investigaciones y creencias que indican que los estudiantes universitarios aprenden bien bajo el régimen de lo que se ha llamado *aprendizaje de dominio* (*mastery learning*). De manera sintética, diremos que el "aprendizaje de dominio" consti-

tuye un sistema pedagógico basado en instrucciones personalizadas que requiere que los estudiantes aprendan unidades curriculares hasta alcanzar un criterio de dominio (véase Abbott y Falstrom, 1975; Ross y McBean, 1995; Senemoglu y Fogelman, 1995; Bergin, 1995). Aunque parece existir cierta investigación que apoya la eficacia del aprendizaje de dominio, hay por lo menos un estudio —y es un buen estudio— realizado por Thompson (1980), cuyos resultados indican que los estudiantes a quienes se enseñó con el método de aprendizaje de dominio no fueron mejores que los estudiantes a quienes se enseñó con un enfoque convencional de conferencia, discusión y memorización. Éste es un estudio ejemplar, realizado con controles cuidadosos, durante un largo periodo. El ejemplo que se presenta a continuación estuvo inspirado en el estudio de Thompson. Sin embargo, el diseño y los controles del ejemplo son mucho más simples que los de Thompson. Observe también que Thompson tenía una enorme ventaja: realizó su experimento en un establecimiento militar, lo cual, por supuesto, significa que muchos problemas de control, con frecuencia recalcitrantes en la investigación educativa, se resolvieron fácilmente.

La controversia surge porque los partidarios del aprendizaje de dominio parecen estar fuertemente convencidos de sus virtudes; mientras que los escépticos permanecen incrédulos. ¿Decidirá la investigación el asunto? Es difícil. Pero ahora se verá cómo se podría diseñar un estudio relativamente modesto, capaz de proporcionar por lo menos una respuesta empírica parcial.

Un investigador educativo decide probar la hipótesis de que el aprovechamiento en ciencia sufre un mayor incremento con un método de aprendizaje de dominio (AD), que con un método tradicional (T). Se ignoran los detalles de los métodos para concentrarse en el diseño de la investigación. Llámese al método de aprendizaje de dominio A_1 y al método tradicional A_2 . Los investigadores saben que otras posibles variables independientes ejercen influencia sobre el aprovechamiento: inteligencia, género, antecedentes de clase social, experiencias previas con la ciencia, motivación, etcétera. Existen razones para creer que los dos métodos funcionan de diferente manera con diferentes tipos de estudiantes. Por ejemplo, quizá funcionen de manera diferente con estudiantes con distintos niveles de aptitud escolar. El enfoque tradicional tal vez resulte efectivo con estudiantes con alta aptitud; mientras que el aprendizaje de dominio sea más efectivo con estudiantes con baja aptitud. Llámese B a las aptitudes: aptitud alta es B_1 y aptitud baja es B_2 . En este ejemplo la variable aptitud se dicotomizó en los grupos de aptitud alta y baja. Ésta no es la mejor forma de utilizar la variable aptitud; cuando una medida continua se dicotomiza o tricotomiza, se pierde la varianza. En un capítulo posterior se verá que constituye un mejor método respetar el nivel de la medida continua y utilizar una regresión múltiple.

¿Qué tipo de diseño debe establecerse? Para responder es importante etiquetar las variables y saber con claridad cuáles son las preguntas que se formulan. Las variables son:

Variable independiente		Variable dependiente
Métodos	Aptitud	Aprovechamiento en ciencias
Aprendizaje de dominio, A_1	Aptitud alta, B_1	Puntuaciones de la prueba de ciencia
Tradicional, A_2	Aptitud baja, B_2	

Los investigadores pudieron haber incluido otras variables en el diseño, en especial variables potencialmente influyentes sobre el aprovechamiento: inteligencia general, clase social, género, promedio de preparatoria, por ejemplo. También se podría utilizar la asignación aleatoria para ocuparse de la inteligencia y otras posibles variables independientes de influencia. La medida de la variable dependiente se puede obtener mediante una prueba estandarizada de conocimientos de ciencia.

FIGURA 18.5

		Métodos		
		A_1 (Aprendizaje de dominio)	A_2 (Tradicional)	
Aptitud	B_1 (Aptitud alta)	$M_{A_1B_1}$	$M_{A_2B_1}$	M_{B_1}
		Puntuaciones en conocimiento científico		
	B_2 (Aptitud baja)	$M_{A_1B_2}$	$M_{A_2B_2}$	M_{B_2}
		M_{A_1}	M_{A_2}	

Parece que el problema requiere de un diseño factorial. Existen dos razones para esta opción: 1) hay dos variables independientes, 2) es claro que se tiene en mente una hipótesis de interacción, aunque no se haya expresado con tantas palabras. Se cree que los métodos funcionarán de manera diferente con distintos tipos de estudiantes. Se establece la estructura de diseño que se representa en la figura 18.5.

Observe que todas las medias marginales y de casilla han sido etiquetadas de forma apropiada. Note también que hay una *variable activa*, métodos; y una *variable atributo*, aptitudes. Quizá recuerde del capítulo 3 que una *variable activa* es una variable experimental o manipulada; una *variable atributo* es una variable medida o una variable que es una característica de personas o grupos; por ejemplo, inteligencia, clase social y ocupación (gente); así como cohesión, productividad y atmósfera restrictiva-permisiva (organizaciones, grupos, etcétera). Todo lo que puede hacerse es categorizar a los participantes como con aptitud alta y aptitud baja, y asignarlos de acuerdo con ello a B_1 y B_2 . Sin embargo, es posible asignar a los estudiantes aleatoriamente a A_1 y A_2 , los grupos de los métodos. Esto se realiza en dos etapas: 1) los estudiantes de B_1 (aptitud alta) se asignan aleatoriamente a A_1 y A_2 , y 2) los estudiantes de B_2 (aptitud baja) se asignan aleatoriamente a A_1 y A_2 . Al aleatorizar así a los participantes se puede suponer que antes de que empiece el experimento, los estudiantes en A_1 son aproximadamente iguales a los estudiantes en A_2 , en todas las características posibles.

El interés aquí radica en los diferentes papeles de la varianza en el diseño de investigación y en el principio de la varianza. Antes de continuar, al principio de la varianza se le llamará “maxmincon” para su fácil referencia. El origen del nombre es evidente: maximizar la varianza sistemática en estudio; controlar la varianza sistemática extraña y minimizar la varianza del error, con dos sílabas invertidas por eufonía.

Antes de ilustrar la aplicación del principio maxmincon en el presente ejemplo, debe discutirse un punto importante. Siempre que se hable de varianza hay que estar seguro de saber de qué tipo de varianza se habla. Se habla de la varianza de los métodos, de inteligencia, de género, de tipo de hogar, etcétera; parecería que se refiere a la varianza de la variable independiente, lo cual es verdad y, a la vez, no. Siempre se refiere a la *varianza de la variable dependiente*, y a la *varianza de las medidas de la variable dependiente*, después de que se realizó el experimento. Esto no es cierto en los llamados estudios correlacionales, donde al decir “la varianza de la variable independiente” significa justamente eso. Al correlacionar dos variables, se estudian “directamente” las varianzas de las variables dependiente e independiente. La referencia “varianza de la variable independiente” surge del hecho de que,

mediante la manipulación y el control de las variables independientes, presumiblemente se ejerce influencia sobre la varianza de la variable dependiente. Dicho de manera algo imprecisa, se "hace" que las medidas de la variable dependiente se comporten o varíen como un supuesto resultado de la manipulación y el control de las variables independientes. En un experimento se analizan las medidas de la variable dependiente y, a partir del análisis, se infiere que las varianzas presentes en la varianza total de las medidas de la variable dependiente se deben a la manipulación y control de las variables independientes y no al error. Ahora regresemos al principio en cuestión.

Maximización de la varianza experimental

La preocupación más obvia del investigador, aunque no necesariamente la más importante, consiste en maximizar la llamada *varianza experimental*. Dicho término se introduce para facilitar discusiones subsecuentes y, en general, tan sólo se refiere a la varianza de la variable dependiente, debida a la influencia ejercida por la variable independiente o variables de la hipótesis sustantiva. En este caso en particular, la varianza experimental es la varianza en la variable dependiente, presumiblemente debida a los métodos A_1 y A_2 , y a los niveles de aptitud B_1 y B_2 . Aunque la varianza experimental puede tomarse para hacer referencia únicamente a la varianza debida a la variable manipulada o *activa*, como los métodos, también se pueden considerar las variables *atributo* como inteligencia, género y, en este caso, aptitud, como variables experimentales. Una de las principales tareas de un experimentador consiste en maximizar esta varianza. Los métodos deben "separarse" lo más posible para hacer a A_1 y A_2 (y A_3 , A_4 , etcétera, si están en el diseño) tan diferentes como sea posible.

Si la variable independiente no varía de manera sustancial, hay poca posibilidad de separar su efecto de la varianza total de la variable dependiente. Es necesario dar a la varianza de una relación la oportunidad de mostrarse, de separarse, por así decirlo, de la varianza total, la cual es un compuesto de varianzas debidas a numerosas fuentes y al azar. Teniendo presente este subprincipio del principio maximin, se puede declarar un precepto de investigación: *diseñar, planear y conducir la investigación de tal forma que las condiciones experimentales sean tan diferentes como sea posible*. Existen, por supuesto, excepciones a este subprincipio, pero probablemente sean poco comunes. Puede ser que un investigador desee estudiar los efectos de pequeñas gradaciones de, digamos, incentivos motivacionales sobre el aprendizaje de algún tema. Aquí no se buscaría que las condiciones experimentales fueran lo más diferentes posibles; aun así, debería asegurarse de que varían un poco o no habría una varianza resultante discernible en la variable dependiente.

En el presente ejemplo de investigación, este subprincipio se refiere a que el investigador debe realizar un esfuerzo para hacer a A_1 y A_2 , los métodos de aprendizaje de dominio y el tradicional, tan diferentes como sea posible. Después, B_1 y B_2 también deben ser tan diferentes como sea posible, en la dimensión de aptitud. Este último problema en esencia es uno de medición, como se verá en un capítulo posterior. En un experimento, el investigador es como un titiritero que hace que los títeres de la variable independiente hagan lo que él quiere. Sostiene los hilos de los títeres A_1 y A_2 con la mano derecha; y los hilos de los títeres B_1 y B_2 , con la mano izquierda. (Se supone que una mano no ejerce influencia sobre la otra, es decir, las manos deben ser independientes.) Los títeres A_1 y A_2 se ponen a bailar por separado, al igual que los títeres B_1 y B_2 ; entonces, el investigador presta atención a la audiencia (la variable dependiente) para observar y medir el efecto de las manipulaciones. Si tiene éxito al hacer bailar a A_1 y A_2 por separado, y si existe una relación entre A y la variable dependiente, y, si por ejemplo, separar A_1 de A_2 es gracioso,

la reacción de la audiencia debería ser una carcajada. El investigador incluso puede notar que sólo consigue risas cuando A_1 y A_2 bailan por separado y, al mismo tiempo, B_1 y B_2 bailan por separado (interacción de nuevo).

Control de variables extrañas

El control de variables extrañas se refiere a minimizar, anular o aislar las influencias de aquellas variables independientes extrañas a los propósitos del estudio. Hay tres formas de controlar las variables extrañas. El primero es el más sencillo, si es posible realizarlo: eliminar la variable como tal. Si existe preocupación sobre la inteligencia como un posible factor contribuyente en estudios de aprovechamiento, su efecto sobre la variable dependiente virtualmente puede ser eliminado utilizando participantes con un solo nivel de inteligencia, digamos puntuaciones de inteligencia dentro del rango de 90 a 110. Si se estudia el aprovechamiento, y el origen racial es un posible factor contribuyente a la varianza del aprovechamiento, se elimina utilizando únicamente miembros de una raza. El principio es: *eliminar el efecto de una variable independiente que posiblemente influya sobre la variable dependiente, es decir, elegir a los participantes de manera que sean lo más homogéneos posible en esa variable independiente.*

Este método para controlar la varianza indeseable o extraña es muy efectivo. Si se selecciona solamente un género para un experimento, entonces se tiene la seguridad de que el género no sea una variable independiente contribuyente. Pero entonces se pierde el poder de generalización; por ejemplo, no es factible hablar sobre la relación estudiada respecto a las niñas si únicamente se utilizan niños en el experimento. Si se restringe el rango de inteligencia, entonces solamente se analiza dicho rango restringido. ¿Es posible que la relación, si se descubre una, sea inexistente o muy distinta con niños de alta inteligencia o con niños de baja inteligencia? Simplemente no se sabe; tan sólo se puede conjeturar o suponer.

La segunda forma para controlar la varianza extraña es a través de la aleatorización. Ésta es la mejor manera, en el sentido de que es posible tener el pastel y también comer un poco de él. En teoría, la aleatorización es el único método para controlar todas las variables extrañas posibles. Otra forma de expresarlo es: si se logra una aleatorización adecuada, entonces los grupos experimentales pueden ser considerados estadísticamente iguales en todas las formas posibles. Por supuesto que esto no quiere decir que los grupos sean iguales en todas las variables posibles. Ya se sabe que, por azar, los grupos pueden ser desiguales; pero con la aleatorización adecuada, la probabilidad de que sean iguales es mayor que la probabilidad de que no lo sean. Por tal razón, el control de la varianza extraña por medio de la aleatorización es un poderoso método de control. Todos los otros métodos dejan abiertas muchas posibilidades de desigualdad. Si se aparean los grupos respecto a la inteligencia, quizá se logre exitosamente la igualdad estadística en inteligencia (al menos en aquellos aspectos de inteligencia que se miden), pero se puede sufrir de desigualdad en otras variables independientes significativamente influyentes como aptitud, motivación y clase social. Un precepto que surge a partir de este poder igualador de la aleatorización es: *siempre que sea posible hacerlo, asigne a los sujetos a los grupos y a las condiciones experimentales de manera aleatoria, y asigne las condiciones y otros factores a los grupos experimentales de manera aleatoria.*

El tercer método para controlar una variable extraña es incluirla en el diseño como una variable independiente. Por ejemplo, suponga que en el experimento discutido anteriormente se fuera a controlar el género y que se considerara inoportuno o imprudente eliminarlo. Se podría añadir una tercera variable al diseño: el género. A menos que se estuviera interesado en la diferencia real entre los géneros respecto a la variable depen-

diente, o se deseara estudiar la interacción entre una o dos de las otras variables y el género, es poco probable que se utilice esta forma de control. Se puede desear información del tipo antes mencionado y también desear controlar el género. En tal caso sería deseable añadirlo al diseño como una variable. El punto es que incorporar una variable a un diseño experimental "controla" la variable, ya que, entonces, resulta posible extraer de la varianza total de la variable dependiente, la varianza debida a la variable. (En el caso anterior se trataría de la varianza "entre géneros".)

Tales consideraciones llevan a otro principio: *una variable extraña puede ser controlada al incorporarla al diseño de investigación como una variable atributo, logrando así control y proporcionando información adicional de investigación sobre el efecto de la variable sobre la variable dependiente y sobre su posible interacción con otras variables independientes.*

La cuarta forma para controlar la varianza extraña consiste en aparear a los participantes. El principio de control detrás del apareamiento es el mismo que aquel para cualquier otra forma de control: el control de la varianza. El apareamiento es similar —de hecho podría llamarse un corolario— al principio del control de la varianza de una variable extraña mediante su incorporación al diseño. El principio básico consiste en dividir una variable en dos o más partes en un diseño factorial, digamos en inteligencia alta y baja, y después aleatorizarla dentro de cada nivel, como se describió antes. El apareamiento es un caso especial de este principio. Sin embargo, en lugar de dividir a los participantes en dos, tres o cuatro partes, se dividen en $N/2$ partes; donde N es el número de participantes involucrados; de esta manera se incorpora al diseño el control de la varianza.

Con el uso del método de apareamiento pueden surgir varios problemas. En primer lugar, la variable respecto a la cual se aparean los participantes debe estar sustancialmente relacionada con la variable dependiente o el apareamiento es una pérdida de tiempo; aun peor, puede causar confusión. Además, el apareamiento tiene limitaciones severas. Si se intenta aparear, digamos, más de dos variables, o incluso más de una, se pierden participantes. Es difícil encontrar participantes apareados en más de dos variables. Por ejemplo, si se decide aparear a los sujetos en cuanto a inteligencia, género y clase social, se puede tener éxito al aparear las dos primeras variables, pero se puede fracasar al intentar encontrar pares que sean bastante iguales en las tres variables. Añádase una cuarta variable y el problema se torna difícil, incluso con frecuencia imposible de resolver.

Sin embargo, no se tire al bebé con el agua del baño. Cuando existe una correlación sustancial entre la variable o variables apareadas y la variable dependiente (mayor que .50 o .60), entonces el apareamiento reduce el término del error y aumenta la precisión de un experimento, un resultado deseable. Si se utilizan los mismos participantes con diferentes tratamientos experimentales —llamados medidas repetidas o diseño de bloque aleatorizado—, entonces se tienen poderosos controles de la varianza. ¿Cómo se puede realizar un mejor apareamiento en todas las variables posibles que apareando un sujeto consigo mismo? Por desgracia, otras consideraciones negativas generalmente impiden dicha posibilidad. Debe enfatizarse con vigor que el apareamiento de cualquier tipo no sustituye la aleatorización. Si se aparean a los participantes, *entonces ellos deben ser asignados aleatoriamente a los grupos experimentales*. A través de un procedimiento aleatorio, como lanzar una moneda o utilizar números pares o nones aleatorios, los miembros de los pares apareados son asignados a los grupos experimental y control. Si los mismos participantes son sometidos a todos los tratamientos, entonces el orden de los tratamientos debe asignarse aleatoriamente. Esto añade control de aleatorización al apareamiento, o control de medidas repetidas.

Un principio sugerido por el presente análisis es: *cuando una variable apareada se correlaciona sustancialmente con la variable dependiente, el apareamiento, como forma de control de la varianza, resulta útil y deseable*. Sin embargo, antes de realizar un apareamiento se

deben sopesar cuidadosamente sus ventajas y desventajas en una situación de investigación particular. La aleatorización completa o el análisis de covarianza pueden ser mejores métodos de control de la varianza.

Otra forma de control, el control estadístico, se analizó en capítulos previos; pero uno o dos comentarios son pertinentes en este momento. Los métodos estadísticos constituyen, por así decirlo, formas de control en el sentido de que aíslan y cuantifican las varianzas. Pero el control estadístico es inseparable de otras formas de control de diseño. Por ejemplo, si se utiliza el apareamiento, debe usarse una prueba estadística apropiada, de otro modo se perderá el efecto del apareamiento y, por lo tanto, el control.

Minimización de la varianza del error

La *varianza del error* es la variabilidad de las medidas debidas a fluctuaciones aleatorias, cuya característica básica es que son *autocompensatorias*, es decir, en un momento varían de una forma, luego de otra, a veces positiva, a veces negativa, a veces hacia arriba, a veces hacia abajo. Los errores aleatorios tienden a equilibrarse entre sí, de tal manera que su media es cero.

Existen varios determinantes de la varianza del error, por ejemplo, factores asociados con las diferencias individuales entre los participantes. Por lo común, a esta varianza debida a diferencias individuales se le llama "varianza sistemática". Pero cuando dicha varianza no puede identificarse ni controlarse, debe agruparse con la varianza del error. Puesto que muchos determinantes interactúan y tienden a cancelarse entre sí (o al menos se supone que lo hacen), la varianza del error tiene estas características aleatorias.

Otra fuente de la varianza del error es aquella asociada con los llamados errores de medición: variación de las respuestas de un ensayo a otro, adivinación, inatención momentánea, fatiga ligera temporal, fallas de la memoria, estados emocionales transitorios de los participantes, etcétera.

Minimizar la varianza del error tiene dos aspectos principales: 1) la reducción de los errores de medición a través de condiciones controladas y 2) un aumento en la confiabilidad de las medidas. Mientras menor sea el control de las condiciones de un experimento, mayor será la influencia de los muchos determinantes de la varianza del error. Ésta es una de las razones para establecer con cuidado condiciones experimentales controladas. En estudios bajo condiciones de campo, por supuesto, dicho control se vuelve difícil; aun así, deben realizarse esfuerzos constantes para disminuir los efectos de los múltiples determinantes de la varianza del error. Esto puede hacerse, en parte, dando instrucciones claras y específicas a los participantes, y excluyendo de la situación experimental los factores que sean extraños al propósito de la investigación.

Incrementar la confiabilidad de las medidas implica reducir la varianza del error. Aunque en posteriores capítulos se incluyen análisis más completos sobre el tema, de momento diremos que la confiabilidad se considera como la precisión de un conjunto de puntuaciones. Hasta el punto en que las puntuaciones no fluctúen aleatoriamente, son confiables. Imagine un instrumento de medición no confiable por completo; dicho instrumento no permite predecir el desempeño futuro de los individuos; en un momento brinda un ordenamiento de los rangos para una muestra de participantes y un ordenamiento diferente de los rangos en otro momento. Con un instrumento de este tipo no sería posible identificar ni extraer las varianzas sistemáticas, debido a que las puntuaciones generadas por el instrumento serían como números en una tabla de números aleatorios, que es el caso extremo. Ahora imaginen cantidades diferentes de confiabilidad y de no confiabilidad en las medidas de la variable dependiente. Cuanto más confiables sean las medidas, mejor se

podrán identificar y extraer las varianzas sistemáticas y menor será la varianza del error en relación con la varianza total.

Otra razón para reducir la varianza del error tanto como sea posible, es darle la oportunidad a la varianza sistemática para que se muestre a sí misma, lo cual no puede hacerse si la varianza del error y, por lo tanto, el término del error, son demasiado grandes. Si existe una relación, se busca descubrirla. Una forma de descubrir la relación consiste en encontrar diferencias significativas entre las medias. Pero si la varianza del error es relativamente grande debido a errores de medición no controlados, la varianza sistemática—llamada antes varianza “entre”—no tendrá la oportunidad de aparecer. Por lo tanto, aunque existente, la relación probablemente no será detectada.

El problema de la varianza del error puede expresarse con claridad en forma matemática: recuerde la siguiente ecuación:

$$V_t = V_e + V_r$$

donde V_t es la varianza total en un conjunto de medidas; V_e es la varianza entre grupos, la varianza presumiblemente debida a la influencia de las variables experimentales; y V_r es la varianza del error (en el análisis de varianza, la varianza dentro de grupos y la varianza residual). En efecto, a mayor valor de V_e , menor deberá ser V_r , con una cantidad dada de V_t .

Considere la siguiente ecuación: $F = V_e/V_r$. Para que el numerador de la fracción a la derecha sea evaluado con precisión respecto a una desviación significativa de las expectativas por el azar, el denominador debe ser una medida exacta del error aleatorio.

Un ejemplo familiar sirve para aclarar esto. Recuerde que en la discusión sobre el análisis de varianza factorial y sobre el análisis de varianza de grupos correlacionados, se habló sobre la varianza debida a las diferencias individuales presentes en las medidas experimentales. Se afirmó que, mientras que la aleatorización adecuada puede igualar efectivamente a los grupos experimentales, habrá varianza en las puntuaciones debida a las diferencias individuales; por ejemplo, diferencias debidas a la inteligencia, aptitud, etcétera. Ahora, en algunas situaciones, estas diferencias individuales pueden ser bastante grandes. Si lo son, entonces la varianza del error y, en consecuencia, el denominador de la ecuación F anterior, serán “demasiado grandes” en relación con el numerador; es decir, las diferencias individuales habrán sido aleatoriamente dispersadas entre, digamos, dos, tres o cuatro grupos experimentales. Aun así, son fuentes de varianza y, como tales, inflarán la varianza dentro de los grupos o la residual, es decir, el denominador de la ecuación anterior.

RESUMEN DEL CAPÍTULO

1. Los diseños de investigación son planes y estructuras utilizados para responder preguntas de investigación.
2. Los diseños de investigación tienen dos propósitos básicos: (i) proporcionar respuestas a preguntas de investigación, y (ii) controlar la varianza.
3. Los diseños de investigación funcionan en conjunto con las hipótesis de investigación para generar una respuesta confiable y válida.
4. Los diseños de investigación también pueden indicar qué prueba estadística emplear para analizar los datos recolectados a partir de ese diseño.
5. Hablar de controlar la varianza, puede referirse a una o más de tres cuestiones:
 - maximizar la varianza sistemática
 - controlar la varianza extraña
 - minimizar varianza del error

6. Para maximizar la varianza sistemática debe tenerse una variable independiente, cuyos niveles sean muy distintos entre sí.
7. Para controlar la varianza extraña, el investigador necesita eliminar los efectos de una variable independiente potencial sobre la variable dependiente, lo cual puede hacerse por medio de:
 - mantener constante la variable independiente; por ejemplo, si se sabe que el género tiene un efecto potencial, entonces puede mantenerse constante al realizar el estudio incluyendo sólo un género (por ejemplo mujeres).
 - la aleatorización, que se refiere a elegir participantes de manera aleatoria y después asignar aleatoriamente a cada grupo de participantes a las condiciones de tratamiento (niveles de la variable independiente).
 - incorporar la variable extraña al diseño, convirtiéndola en una variable independiente.
 - apareando a los participantes; este método de control puede resultar difícil en ciertas situaciones; un investigador nunca podrá estar muy seguro de que se realizó un apareamiento exitoso en todas las variables importantes.
8. Minimizar la varianza del error incluye la medición de la variable dependiente. Al reducir el error de medición se reduce la varianza del error. El incremento en la confiabilidad de la medición también conlleva una reducción de la varianza del error.

SUGERENCIAS DE ESTUDIO

1. Hemos notado que el diseño de investigación tiene el propósito de obtener respuestas a las preguntas de investigación y controlar la varianza. Explique en detalle qué significa esta afirmación. ¿Cómo controla la varianza un diseño de investigación? ¿Por qué un diseño factorial debería controlar más varianza que un diseño de un factor? ¿Cómo es que un diseño que utiliza participantes apareados o medidas repetidas de los mismos participantes controla la varianza? ¿Cuál es la relación entre las preguntas de investigación, las hipótesis de investigación y un diseño de investigación? Invente un problema de investigación para ilustrar sus respuestas a estas preguntas (o utilice un ejemplo del texto).
2. Sir Ronald Fisher (1951), el inventor del análisis de varianza, en uno de sus libros dijo que debe aclararse que la hipótesis nula nunca se confirma o establece; pero que es posible refutarla en el curso de la experimentación. Se puede decir que todo experimento existe para dar a los hechos la oportunidad de refutar la hipótesis nula. Ya sea que usted esté de acuerdo o no con la afirmación de Fisher, ¿qué piensa que quiso decir con ello? Para estructurar su respuesta, recuerde el principio *maximincon* y las pruebas *F* y *t*.