

1.2. CUATRO TIPOS DE VARIABLES

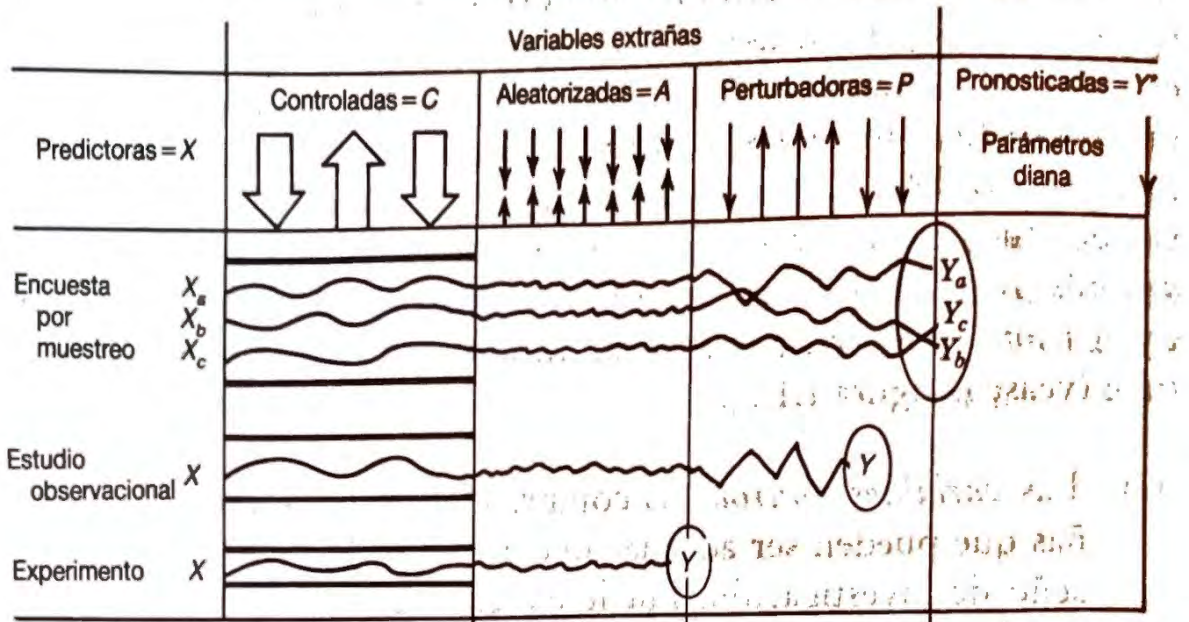
(E) Las *variables explicativas* constituyen el objetivo del diseño de la investigación. Se las denomina a veces variables *experimentales*, mientras que el término *explicativas* ha sido utilizado para lo que designaré como variables predictoras. Aquí *explicativas* se refiere a aquellas variables que expresan los objetivos del diseño de la investigación, y entre las cuales el investigador desea encontrar y medir alguna relación preestablecida. Las variables explicativas comprenden dos grupos diferenciados: las *predictoras* (X) incluyen las causas buscadas de las relaciones, y las *pronosticadas* (Y) describen los efectos de la predicción. Se han utilizado otros nombres para estos dos grupos de variables: independientes y dependientes, estímulo y respuesta, tratamiento y criterio, causa y efecto, X e Y , así como otras parejas de términos análogos. También se ha utilizado la expresión determinantes y consecuencias, especialmente en la investigación demográfica. No nos sentimos obligados a explorar aquí las sutiles distinciones filosóficas entre esos pares de términos. La terminología carece de uniformidad y optaré por términos neutrales. Además, en los estudios observacionales y en las encuestas, la división de las variables explicativas en predictoras y pronosticadas puede resultar más o menos arbitraria.

Se definen primero las variables explicativas que incluyen los objetivos de investigación, sobre la base de teorías científicas sustanti-

vas. Estas variables surgen del conocimiento y la comprensión del campo de estudio. Es inevitable, sin embargo, reconocer también la existencia potencial de otras fuentes *extrañas* de variación, razón por la que es preciso idear métodos que permitan separar estas variables extrañas de las variables explicativas. Nos parece una simplificación útil el clasificar todas las diferentes variables extrañas en tres tipos. Además, no debe surgir confusión al hablar aquí de “variables”, en vez de “fuentes” de variación. *Las variables explicativas (E), las predictoras (X) y las pronosticadas (Y) materializan los objetivos de la investigación, y los otros tres tipos de variables son extraños a esos objetivos* (véase la figura 1.1).

- (C) Las *variables controladas* comprenden aquellas variables extrañas que pueden ser adecuadamente controladas mediante el diseño de investigación. Puede ejercerse el control bien a través del diseño de los procedimientos de selección, o bien mediante técnicas de estimación en el análisis estadístico, o quizás simultaneando ambos procedimientos. Las opciones dependerán de la previsión y del conocimiento, pero también de la disponibilidad y uso estratégico de datos y recursos. Las técnicas para controlar las variables extrañas tienen por objeto disminuir los errores aleatorios (tipo A), o disminuir los efectos de sesgo de las variables perturbadoras (tipo P), o ambos.
- (P) Las *variables perturbadoras* son aquellas variables extrañas no controladas que pueden ser confundidas con las variables explicativas (tipo E). La principal desventaja y preocupación de los diseños no experimentales reside en el fallo en trasladar todas estas variables P bien al tipo C de variables controladas, o bien al tipo A de variables aleatorizadas.
- (A) Las *variables aleatorizadas* son variables extrañas no controladas que son tratadas como errores aleatorios. En los experimentos “ideales” se hace que sean de hecho y operacionalmente aleatorizadas, pero en las encuestas y en otras investigaciones solamente se supone que están aleatorizadas, según se discutirá en la sección 1.4. Puede considerarse la aleatorización como una forma de control experimental, aunque distinta de las formas utilizadas para las variables de tipo C.

FIGURA 1.1. Efectos de tres tipos (C, A, P) de variables extrañas sobre las variables explicativas (E) ($X \rightarrow Y$)



Tres tipos de variables extrañas pueden obstaculizar el paso de las variables explicativas (E), desde las predictoras (X) a las pronosticadas (Y). Algunas variables son controladas (C) mediante diversos tipos de controles (bloques, estratificación, apareamiento, ponderación, etc.). Otras son aleatorizadas (A) en los experimentos, o tratadas como aleatorizadas en las encuestas y en los estudios observacionales. Sin embargo, algunos efectos permanecen como variables de perturbación (P) en las encuestas y en los estudios observacionales, pero ninguno de estos efectos permanece en los "verdaderos" experimentos.

Las flechas hacia arriba y hacia abajo representan los efectos contrapuestos de las variables extrañas. Las variables controladas aparecen con trazo más grueso que las variables perturbadoras, porque se confía en que las variables más fuertes sean controladas por el diseño. Las variables aleatorizadas se trazan como numerosas, delgadas y actuando en ambas direcciones, tendiendo a neutralizarse entre ellas; sus efectos sobre el curso ($X \rightarrow Y$) se muestran en la figura como relativamente débiles.

Los experimentos son los que aparecen controlados más estrechamente; y el muestreo de encuestas lo es mucho menos, si bien son capaces de representar bases de población más amplias, que cubren varios dominios, como lo muestran los tres caminos ($X \rightarrow Y$), a, b, c.

Sin embargo, las variables pronosticadas Y de los experimentos se muestran mucho más alejadas del parámetro objetivo Y*, que representa el efecto/respuesta final que el estudio trata de predecir/estimar. Estos espacios vacíos aluden a la relativa falta de "realismo" de los diseños del estudio.

Nuestro objetivo consiste en colocar dentro del tipo C (variables controladas), mediante diseños *eficientes*, tantas variables extrañas como sea posible, práctico y económico. Sin embargo, no es posible o práctico controlar más que unas pocas de las potenciales variables perturbadoras y algunas o la mayoría de ellas deben quedar sin controlar. El objetivo de la aleatorización en los experimentos reside en colocar todas las variables de tipo P dentro de las de tipo A. En los experimentos "ideales" no se dejan variables dentro del tipo P. Todas las variables extrañas han sido, o bien controladas dentro del tipo C, o bien aleatorizadas dentro del tipo A. Al colocar las variables perturbadoras en el tipo C, eliminamos los efectos que tendrían que estar en el tipo A. Si bien los errores aleatorios quedan en el tipo A, en los experimentos "ideales" se eliminan los sesgos debidos al tipo P.

Sin embargo, en la investigación no experimental —encuestas y otras investigaciones— los controles tienen que cumplir una doble función. Incrementan la eficiencia al disminuir los errores de las variables de tipo A, tal como lo harían en los experimentos reales. Pero sin aleatorización no podemos eliminar completamente todas las variables perturbadoras de tipo P. En la medida en que estas últimas queden sin estar completamente controladas dentro del tipo C, o sin haber sido aleatorizadas dentro del tipo A, las variables de tipo P quedan mezcladas y confundidas con las variables explicativas de tipo E, ejerciendo sobre ellas efectos de sesgo desconocidos. Por tanto, la reducción de esos sesgos se convierte en la función crucial de los controles en la investigación no experimental.

1.3. ENCUESTAS, EXPERIMENTOS E INVESTIGACIONES CONTROLADAS

Se han desarrollado los diseños experimentales para verificar y determinar las variables explicativas y para medir las relaciones entre ellas en los contrastes analíticos con los datos. De manera distinta y diferenciada, la teoría del muestreo de encuestas se ha desarrollado principalmente para aportar estadísticas descriptivas, medias, proporciones y totales, especialmente para grandes muestras extraídas de poblaciones mucho más grandes. Sin embargo, los datos de las encuestas se utilizan frecuentemente y con éxito, especialmente en las ciencias sociales, para la investigación estadística y analítica de las re-

laciones entre variables explicativas. Además, en muchas situaciones, ni los experimentos ni las encuestas por muestreo resultan viables o prácticas, por lo que muchas veces debe utilizarse algún tipo de investigación controlada o de estudio observacional. El desarrollo teórico de cada uno de estos tres tipos de diseño ha acontecido dentro de un espléndido aislamiento, con el fin de simplificarlos y clarificarlos. No se han estudiado bien las conexiones y contrastes entre ellos, pero estamos a punto de emprender su exploración.

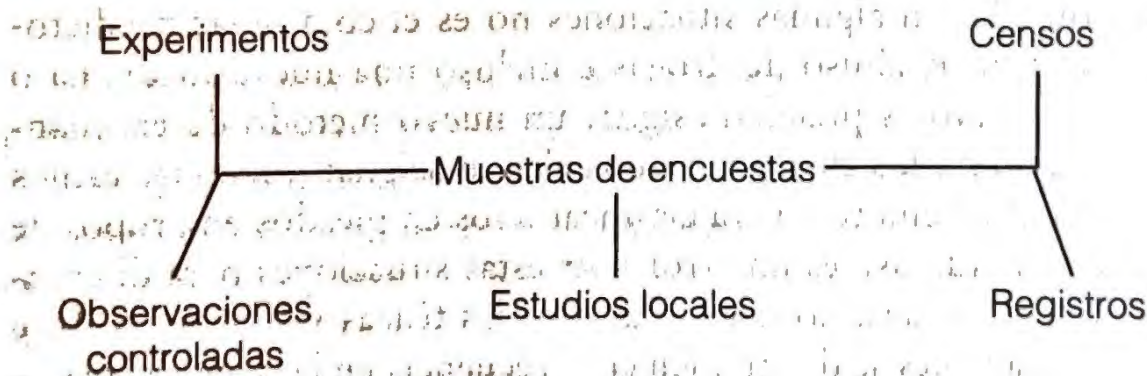
Con el término *experimentos* quiero referirme aquí a los experimentos "ideales", en los que todas las variables extrañas han sido o bien controladas, o bien aleatorizadas (1.4). Al hablar de *encuestas* (o *encuestas por muestreo*), me refiero a muestras probabilísticas, en las que todos los elementos de una determinada población tienen una probabilidad positiva conocida de ser seleccionados en la muestra (1.7). Entiendo por investigaciones (o *investigaciones controladas*), la recogida de datos (realizada con esmero, y a veces bajo un considerable control), pero sin la aleatorización de los experimentos, o sin el muestreo probabilístico de las encuestas (capítulo 3). Las diferencias entre experimentos, encuestas e investigaciones no son las consecuencias del análisis estadístico, que pueden ser similares, sino que resultan de los diferentes diseños a la hora de introducir las variables y de seleccionar los elementos de la población (sujetos).

Al tomar en consideración los fines más amplios de cualquier investigación científica, sólo puede controlarse de modo objetivo y sólido una parte de los medios totales (es decir, recursos) requeridos para hacer inferencias. Las demás partes tienen que quedar sujetas al criterio más o menos vago y subjetivo (por más cualificado que sea). Los científicos buscan maximizar la primera parte y, consiguientemente, minimizar la segunda. Al evaluar los fines, los costes y los medios disponibles, efectúan una elección estratégica de métodos. De esta manera se enfrentan con los tres problemas básicos de la investigación científica: aleatorización, representatividad y realismo de las mediciones.

La experimentación es potente en cuanto al control de las variables explicativas por medio de la aleatorización de las variables predictoras en los sujetos (es decir, asignación de los sujetos a tratamientos); pero es débil en cuanto a la representatividad con respecto a las poblaciones objetivo definidas, y muchas veces también en cuanto al realismo de las mediciones. Las encuestas tienen su lado fuerte en la representatividad, y su lado débil en el control de las variables. Las

investigaciones son deficientes en cuanto al control y muchas veces en cuanto a la representatividad. Su gran preponderancia suele obedecer a su bajo coste y relativa comodidad, y a veces a la necesidad y viabilidad del realismo de las mediciones en “entornos naturales”. Habitualmente nos vemos enfrentados a conflictos entre las aspiraciones de aleatorización, de representatividad y de realismo. Raras veces podrán satisfacerse adecuadamente los deseos de cumplir con los tres criterios en un diseño de investigación, y muy a menudo no se podrán satisfacer ni siquiera dos de los tres. Lo más corriente es que simplemente se destaque uno de los criterios porque es el menos costoso y el más práctico, y porque puede parecer —de manera convincente o sólo a nivel de intenciones— el más justificado por motivos teóricos. Pero cuando la esperanza o el deseo son los padres del pensamiento, la criatura puede ser ilegítima. Debe prestarse mucha más atención a los criterios a los que haya que renunciar. Sería deseable otorgar un papel más destacado a modelos explícitos de las diversas fuentes de variación que surgen de los tres tipos de diseños de investigación (véase figura 1.2).

FIGURA 1.2. *Tres comparaciones distintas de las encuestas por muestreo*



En la sección 1.3 se analizan las comparaciones de las encuestas por muestreo con los experimentos y con las observaciones controladas. Estos tres métodos están justificados en calidad de opciones estratégicas preferidas para los tres criterios de representatividad, tratamientos aleatorizados y realismo. Otro tipo de opción estratégica pone en juego las ventajas relativas de las encuestas por muestreo frente a los censos o registros (5.2). Otra comparación más es la relativa a la elección entre encuestas muestrales amplias y estudios locales confinados a una o a pocas áreas (3.1). De esta manera, las encuestas por muestreo se presentan como alternativa a tres métodos claramente diferentes de recogida de datos, y ésta es la única razón de que ocupen un lugar central en la figura.

Los experimentos gozan de tres ventajas principales: 1. Es posible eliminar los sesgos originados por las variables perturbadoras (tipo *P*) mediante la aleatorización de las variables extrañas. 2. Los controles sobre la introducción y variación de las variables predictoras clarifican tanto la dirección como la naturaleza de la causalidad existente entre las variables predictoras y las pronosticadas. Por el contrario, en el caso de las relaciones halladas en resultados de encuestas, no están claras ni la dirección ni la naturaleza de la causalidad. 3. El diseño experimental moderno permite una gran flexibilidad, eficiencia, simetría y una potente manipulación estadística, mientras que el uso analítico de los datos de encuesta presenta difíciles problemas estadísticos (2.7 y 7.1).

Son tan bien conocidas las ventajas del método experimental que no necesitamos hacer aquí hincapié sobre ellas. Es considerado el método científico por excelencia, siempre que sea factible. En muchas situaciones los experimentos no son posibles, y éste es frecuentemente el caso de las ciencias sociales, donde los predictores no pueden asignarse libremente a los sujetos por diversas razones: (1) En la mayoría de los casos es imposible asignar predictores tales como edad, sexo, religión, cociente intelectual, e incluso ingresos y educación. (2) En algunas situaciones no es ético asignar predictores como fumar, abuso de drogas e incluso una nueva medicina o placebo. (3) No es práctico asignar un nuevo método de enseñanza para dividir a los alumnos de una clase o asignar diferentes estilos de dirección de empresa para clasificar a los empleados en grupos de trabajo, etc. Aun así, es un error usar estas situaciones para diferenciar entre las ciencias sociales y las ciencias físicas y biológicas. Este tipo de situaciones se da también con frecuencia en las ciencias físicas (en meteorología, astronomía, geología), en las ciencias biológicas y en medicina, ingeniería, industria y en el mundo de los negocios, etcétera (1.4).

Incluso allí donde es aplicable, el método experimental puede presentar también inconvenientes que deben o bien superarse o bien tolerarse. En primer lugar, a menudo puede resultar difícil elegir y manejar las variables de control; puede ser difícil o imposible diseñar un experimento "ideal". Los periódicos publican frecuentemente informes acerca de éxitos con nuevos medicamentos y curaciones que luego pueden desmentirse o caer en el olvido. Discutiremos más adelante (capítulo 4) la necesidad y las dificultades asociadas al establecimiento de controles apropiados, y los lectores tienen seguramente

sus propias ideas y ejemplos al respecto. Nos limitamos a mencionar aquí la utilización de placebos para ocultar los tratamientos a los sujetos, y el uso de "ensayos clínicos doble ciego" para ocultar los tratamientos tanto a los sujetos como a los investigadores (1.4). Nos referiremos a los famosos experimentos Hawthorne con trabajadores (3.6) y al gran Pavlov con sus perros.

En los experimentos reales el "ideal" no se alcanza fácilmente, y por lo tanto la distinción global entre la investigación experimental y la no experimental no es absoluta. Los problemas con los controles experimentales llevaron incluso al gran Pavlov a creer, equivocadamente, por algún tiempo, que había probado el carácter hereditario de una habilidad adquirida para aprender: «En una declaración informal, hecha durante el XIII Congreso Internacional de Fisiología en Boston, en agosto de 1929, Pavlov explicó que al revisar estos experimentos se encontró que la mejora aparente en la habilidad para aprender de generaciones sucesivas de ratones se debía, en realidad, a una mejora en la habilidad para enseñar por parte del experimentador» [Greenberg, 1969, p. 327]. Sin embargo, la distinción es real y útil. Por tanto, sir Austin Bradford Hill exagera su acertado argumento al decir: «las dificultades de los experimentos no son menores» cuando discutía las dificultades de los estudios observacionales [en Cochran, 1965]. [Kish, 1959, 1975.]

Así pues, las ventajas de los experimentos sobre las encuestas, al permitir un mejor control de los tratamientos, son sólo relativas y no absolutas. El diseño de controles experimentales apropiados no es automático, sino que se trata de un arte que requiere conocimiento científico y creatividad a la hora de plantear el experimento, y capacidad retrospectiva al interpretar los resultados. Sin embargo, la distinción entre experimentos y encuestas en cuanto al control es real e importante. Para subrayar esta distinción nos referimos aquí a los experimentos "ideales", en los que el control de los sesgos provenientes de las variables perturbadoras es total.

En segundo lugar, en general es difícil diseñar experimentos de manera tal que sean representativos de una población específica importante. De hecho, los interrogantes y problemas del muestreo, es decir, de la *representatividad* de las poblaciones implicadas en los resultados experimentales, han sido en gran medida ignorados en el diseño experimental. Tanto en la teoría como en la práctica, la investigación experimental ha dejado a menudo de lado la verdad básica de que los sistemas causales, la distribución de relaciones (como las distribuciones de características), existen solamente dentro de universos

concretos. Las inferencias estadísticas derivadas de la contrastación experimental de diversos tratamientos se limitan a la población o poblaciones incluidas en el diseño experimental. No obstante, tienen que recibir aplicaciones mucho más amplias, y para ello debemos recurrir a modelos, bien explícitamente o, demasiado a menudo, sólo de modo implícito. Más adelante trataremos de forma concisa estos controvertidos temas (1.4 y 1.8).

En tercer lugar, para muchos objetivos de investigación, especialmente en las ciencias sociales, no es factible reproducir, en el diseño experimental, el deseado *realismo* de un "entorno natural" de cara a las mediciones. Por eso, los experimentos sociales a veces dan respuesta a preguntas que sólo tienen un significado vago. Es decir, las variables del experimento, construido artificialmente, pueden tener solamente una relación tenue con las variables que el investigador desearía investigar.

Los dos últimos puntos débiles de los experimentos que hemos indicado señalan las ventajas de las encuestas. Las muestras probabilísticas no sólo permiten inferencias estadísticas claras a poblaciones definidas, sino que frecuentemente pueden realizarse las mediciones en los "entornos naturales" de poblaciones reales. De esta manera, en las situaciones concretas de investigación, tanto el método experimental, como el de encuestas, presentan sus problemas e inconvenientes específicos, así como también sus ventajas.

En la investigación social, con frecuencia se eligen y se prefieren algunos diseños de investigaciones controladas y no las encuestas o los experimentos. Esta elección preferencial se hace por razones de coste o de viabilidad o para mantener el deseado realismo de las mediciones. Se han ideado ingeniosas adaptaciones de los diseños experimentales para estas investigaciones controladas (capítulo 3). Se utiliza muchas veces el marco estadístico y el análisis de los diseños experimentales, pero sin la aleatorización de los verdaderos experimentos. A menudo se requiere una gran capacidad de inventiva en estos diseños a fin de lograr flexibilidad, eficiencia y especialmente cierto grado de control de las variables extrañas (capítulo 4). Las investigaciones controladas toman muchas formas y tienen muchos nombres: estudios observacionales, observaciones controladas, cuasi- o pseudoexperimentos, experimentos naturales y otras. Prefiero no tomar prestada la prestigiosa palabra *experimento* para estudios en los que los predictores no están aleatorizados, y prefiero evitar términos como *experimentos ex-post-facto*, que confunden más que ayudan.

Las investigaciones controladas y las encuestas muestrales no son simples experimentos de segunda clase, sino que tienen su propia justificación.

En la práctica, carecemos generalmente de los recursos necesarios para superar todas las dificultades y, por tanto, para conseguir simultáneamente la perfección en el realismo de las medidas, en la aleatorización para controlar los tratamientos y en la representatividad de las grandes poblaciones. Admitamos que a menudo, incluso casi siempre, no podemos satisfacer los tres criterios simultáneamente. Después de admitir todo esto, sin embargo, muchos autores proclaman una jerarquía general entre estos criterios, de manera que uno de ellos debe predominar en todas las situaciones. Por ejemplo, algunos creen que, siempre que sea posible, debe lograrse a cualquier precio el carácter aleatorio de los tratamientos (o "*validez interna*"), antes de prestar atención a la representatividad con respecto a las poblaciones (o "*validez externa*"), o al realismo. Por el contrario, mi punto de vista es que no existe un supercriterio que pueda detentar la superioridad única, omnímoda y omnipresente sobre los demás. Más bien hay que escoger y comprometerse con una estrategia de investigación de modo que nuestros recursos se ajusten a la situación concreta. En cualquier situación específica, un método puede ser mejor o más práctico que los otros, pero no existe la superioridad general en todas las situaciones de ninguno de los tres diseños. La comprensión de los puntos fuertes y débiles de cada uno de ellos debería permitir una elección más acertada. A veces, la flexibilización de uno de los criterios puede acarrear grandes ventajas en otro, o incluso en los dos criterios restantes. Además, algunos problemas de investigación difíciles han de abordarse por separado mediante dos o incluso los tres métodos.

Estas perspectivas antijerárquicas sobre estrategia están parcialmente en desacuerdo con algunos de los mejores escritos sobre el control de variables en los experimentos e investigaciones sociales, realizados por Campbell (1957) y Campbell y Stanley (1963), con los que el lector debería estar familiarizado (véase el final de la sección 3.5).

Puede mejorarse cada uno de los tres tipos de diseño esforzándose en superar sus puntos débiles específicos. Dado que la falta de control sobre los tratamientos es la principal debilidad de las encuestas, los investigadores de encuestas deberían mejorar la recogida y el uso de variables auxiliares como controles de las variables perturba-

doras. Deberían estar más atentos a los cambios sociales y utilizarlos para medir los efectos de los “experimentos naturales”. Deberían, asimismo, explorar más a menudo los datos de encuestas, haciendo uso de técnicas analíticas multivariantes, tendencia que está bastante afianzada en la actualidad.

Por otra parte, pueden mejorarse los experimentos e investigaciones controladas muchas veces esforzándose en delimitar más claramente sus poblaciones y haciendo que sus resultados sean más representativos de dichas poblaciones. A menudo se puede y se debe hacer más para ampliar la base del diseño con objeto de poder extender la inferencia estadística a poblaciones más amplias y significativas. Con demasiada frecuencia y precipitación, los investigadores se comprometen con poblaciones pequeñas, cautivas y cómodas de manejar. Demasiado a menudo el investigador justifica estas restricciones como un intento de lograr que los sujetos sean más “homogéneos”. Si el sentido común no desbarata este error, la lectura de R. A. Fisher lo hará.

Hemos visto que el diseño factorial posee dos ventajas frente a los experimentos en los que intervienen únicamente factores aislados: (i) Mayor *eficiencia*, en el sentido de que estos factores son evaluados con la misma precisión utilizando sólo la cuarta parte de las observaciones que, de otra manera, hubiesen sido necesarias; y (ii) mayor *alcance*, ya que además de los cuatro efectos de cada uno de los factores, se evalúan sus 11 posibles interacciones. Hay una tercera ventaja que, aunque menos obvia que las dos anteriores, influye de manera importante sobre la utilidad de los resultados experimentales en su aplicación práctica. Se trata de que cualquier conclusión, como, por ejemplo, la conveniencia de aumentar la cantidad de un ingrediente dado, tiene una base inductiva más amplia cuando se infiere a partir de un experimento en el que se han alterado las cantidades de los otros ingredientes, de la que tendría con cualquier número de experimentos en los que estos ingredientes se hubieran mantenido rígidamente constantes. La estandarización minuciosa de las condiciones experimentales, que se suele invocar con frecuencia de modo irreflexivo como una panacea, lleva siempre consigo la desventaja real de que un experimento con altos niveles de estandarización aporta información directa sólo con respecto al estrecho rango de las condiciones obtenidas a través de la estandarización. Por consiguiente, la estandarización debilita, en vez de fortalecer, nuestra base para inferir resultados similares en caso de que cambien algo dichas condiciones, como sucede invariablemente en la práctica [Fisher, 1935, p. 99].

El investigador debería contemplar la base poblacional en términos de factores de coste y componentes de variación, y luego ampliar