

capítulo

5

Diseños experimentales: segunda parte

Introducción

El presente capítulo amplía algunos conceptos vertidos en el capítulo 7 de, *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*, en particular, los referidos a las fuentes de invalidación interna, la necesidad de varios grupos de comparación, la equivalencia de los grupos experimentales y la técnica de emparejamiento, así como el diseño experimental de Solomon. Asimismo, se revisan otros diseños experimentales: las series cronológicas múltiples y los factoriales. Finalmente, se comentan los diseños cuasiexperimentales.

Fuentes de invalidación interna (ampliación y ejemplificación)

En el libro impreso se describieron estas fuentes y se recomienda eliminarlas mediante el control, lo que The SAGE glossary of the social and behavioral sciences (2009p) y León y Montero (2003) denominan “descartar causas alternativas o extrañas”. A continuación se amplían y ejemplifican.

1. *Historia*. Estas fuentes constituyen acontecimientos que ocurren durante el desarrollo del experimento, que pueden afectar a la variable dependiente y llegar a confundir los resultados. Por ejemplo, durante un experimento para analizar el efecto que tienen distintas formas de realimentación en el trabajo sobre la motivación (vía el supervisor “cara a cara” y por escrito), se consideran dos grupos de obreros: a uno le aumentan el salario o se le felicita por su desempeño laboral mientras se lleva a cabo el experimento y al otro grupo, no. Las diferencias en la variable dependiente pueden atribuirse a la manipulación de la independiente o al acontecimiento que sucedió durante el experimento. Supongamos que experimentamos la mejora de procesos industriales y el estímulo es la introducción de un sistema ISO de calidad, en una sola planta se

introduce este y en otra no. Para tener certeza respecto de los resultados debemos evitar que en ambas ocurra algo que altere significativamente tal mejora y que sea independiente del estímulo. O bien, imaginemos que en un experimento donde se prueba la eficacia de un medicamento y se tienen grupos pequeños (digamos de cuatro personas cada uno), al grupo experimental se les administra el medicamento y al grupo de control, no. Sin embargo, no sabemos que dos miembros de este último se automedican en secreto, lo cual invalida los resultados del experimento.

2. *Maduración.* Son procesos internos de los participantes que operan como consecuencia del tiempo y afectan los resultados del experimento, tales como cansancio, hambre, aburrimiento, envejecimiento y cuestiones similares. Si a un grupo de niños se le expone a un nuevo método de aprendizaje por dos años, los resultados pueden estar influidos simplemente por la maduración de los infantes durante el tiempo que duró el experimento. En un experimento quizá los sujetos pueden cansarse y sus respuestas serán afectadas por ello. Si tenemos dos grupos y la condición experimental del primero implica mucho más tiempo que la del segundo, puede alterar esta fuente. Lo mismo sucede cuando experimentamos con medicamentos, intervenciones o tratamientos médicos o psicológicos. En estos casos, el factor tiempo tiene que ser idéntico entre los grupos participantes, al igual que cuando experimentamos con sustancias o elementos como el fuego.
3. *Inestabilidad del instrumento de medición.* En este caso, la poca o nula confiabilidad de las mediciones provoca desconfianza en las puntuaciones obtenidas, es decir, no sabemos si las deducciones son certeras porque el instrumento a veces da unos resultados y luego, otros. Este inconveniente se evita mediante el diseño de instrumentos válidos y confiables (vea el capítulo 9 del libro: "Recolección de los datos cuantitativos").
4. *Inestabilidad del ambiente experimental.* Tal como se mencionó en el capítulo 7 del texto impreso, las condiciones o componentes del ambiente o entorno del experimento pueden ser desiguales para los grupos participantes. Imaginemos que dos grupos son sometidos a un mismo experimento sobre memorización y, al sortearlos, a uno le corresponde realizarlo en un aula cerca de donde se efectúa una remodelación del edificio y al otro, en un recinto donde priva el silencio. Si el experimento requiriera de una elevada concentración de los individuos que participan en él, la concentración de un grupo podría ser diferente de la del otro (a causa del ruido, del tránsito de personas, etc.), lo que afectará la interpretación de los resultados. Tal vez las diferencias entre los grupos se deban a variaciones en la concentración y no a la variable independiente, o es posible que esta sí ejerza algún efecto, pero no podemos estar seguros de ello. No tenemos confianza en los resultados. En experimentos en ciencias naturales es claro que cuestiones como la temperatura ambiental deben estar perfectamente controladas. Asimismo, en experimentos médicos el entorno debe ser similar para todos.
5. *Administración de pruebas.* Se refiere al efecto que puede tener la aplicación de una prueba sobre las puntuaciones de pruebas subsecuentes. Por ejemplo, si en un experimento sobre prejuicio a un grupo se le aplica una prueba para detectar su nivel de prejuicio, y luego se le expone a un estímulo experimental (que supuestamente debe

reducir el prejuicio), y después se mide nuevamente el prejuicio para evaluar si disminuyó o no, puede ocurrir que las puntuaciones de prejuicio de la segunda medición (después del estímulo) sean influidas por la aplicación de la primera prueba (antes del estímulo). Es decir, la administración de la primera medición puede sensibilizar a los participantes del experimento y, cuando responden a la segunda prueba, sus respuestas podrían estar afectadas por esa sensibilización. Si disminuye el prejuicio y no hay control, no sabremos cuánto se debió al estímulo experimental o a la variable independiente y cuánto a dicha sensibilización.

6. *Instrumentación*. En este caso se hace referencia a cambios en los instrumentos de medición o en los observadores, modificaciones que pueden producir variaciones en los resultados. Si la prueba del grupo experimental es diferente a la del grupo de control u otro grupo experimental, puede intervenir la instrumentación. Por ejemplo, si queremos ver el efecto de dos métodos de enseñanza, sometemos a cada grupo a un método diferente y después aplicamos un examen de conocimientos para ver la eficacia de cada método y comparar los resultados. Si los exámenes no fueran equivalentes podría presentarse esta fuente. Imaginemos que el examen de un grupo es más fácil que el del otro. En este caso, ¿cómo podremos estar seguros de que las diferencias entre las puntuaciones de los exámenes se deben al estímulo (método de enseñanza) y no a que las mediciones son distintas? Esta incertidumbre se repite en el caso de un experimento para probar la eficacia de dos medicamentos que reducen la hipertensión. El procedimiento y no solo el aparato para medir la presión arterial debe ser el mismo (incluyendo número de mediciones, tiempo entre ellas, posición de los pacientes, etcétera).
7. *Regresión estadística*. Esta fuente se presenta si seleccionamos participantes sobre la base de puntuaciones extremas. y cuando se miden por primera vez se encuentran en valores muy altos o bajos en la variable que nos interesa, después tienden a regresar a su estado normal, y en una segunda medición obtienen valores no extremos; la comparación entre las dos mediciones indica un cambio, pero en realidad lo que ocurre es que los medimos inicialmente cuando su condición en la variable era extrema. En términos de Campbell y Stanley (1966), se trata de un efecto provocado por la tendencia que muestran los participantes seleccionados sobre la base de puntuaciones extremas a regresar en pruebas posteriores a un promedio en la variable en la que fueron medidos.

Por ejemplo, pensemos que pretendemos evaluar el efecto de liderazgo autocrático de un profesor en la ansiedad de los alumnos. Primero, aplicamos al grupo una prueba para medir la ansiedad. Después, exponemos a los participantes a un profesor autocrático y repetimos la medición. Ahora, imaginemos que los estudiantes están muy estresados durante la aplicación de la primera prueba porque tienen un examen difícil al día siguiente. En estas circunstancias, podría ocurrir que se observara que el grupo está “menos ansioso” después de recibir la manipulación, y que, aparentemente, el profesor autocrático redujo la ansiedad.

Pero en realidad lo que sucedió fue que durante la segunda prueba estaban “retornando” a su nivel normal de ansiedad. Otro caso sería el de una medición de la pre-

sión arterial en la preprueba, después de que las personas tuvieron que subir escaleras prolongadas o desplazarse rápidamente para que se les hiciera el examen respectivo u otras mediciones cuando no estaban en ayunas.

La regresión estadística representa el hecho de que puntuaciones extremas en una distribución particular tenderán a desplazarse (esto es, regresar) hacia el promedio de la distribución en función de mediciones repetidas (Christensen, 2006). Entre una primera y una segunda mediciones, las puntuaciones más altas tienden a bajar y las más bajas a aumentar. En ocasiones, este fenómeno de regresión se presenta porque ambas mediciones no están adecuadamente correlacionadas.

8. *Selección.* Esta fuente puede presentarse al elegir a las personas para los grupos del experimento, de tal manera que estos no sean equiparables. Es decir, si no se escogen los participantes de los grupos asegurándose su equivalencia, la selección resulta tendenciosa. Por ejemplo, en un experimento sobre métodos educativos, si en un grupo se incluyen a los estudiantes más inteligentes y estudiosos, y en el otro a los estudiantes menos inteligentes y abúlicos, las diferencias entre los grupos se deberán a una selección sesgada, aparte del tratamiento experimental o variable independiente. Lo mismo sucede con ciertos experimentos médicos: imaginemos que al probar un tratamiento un grupo está conformado básicamente por individuos jóvenes y sanos, y el otro, por personas significativamente mayores. En experimentos con materiales sometidos a distintas condiciones, dichos materiales deben ser completamente equivalentes.
9. *Mortalidad experimental.* Esta situación se refiere a las diferencias entre las pérdidas de participantes que sufren los grupos que se comparan. Si un grupo pierde 25% de sus miembros y el otro grupo solo 2%, los resultados serán influidos por ello, además de por el tratamiento experimental. La pérdida de participantes puede deberse a diversas razones, por ejemplo, desmoralización, fatiga ante el tratamiento o desesperación por no ver progreso (Grinnell et al., 2009), por lo que algunos autores la incluyen como una fuente de invalidación interna por sí misma (Hulley, Cummings, Browner, Grady y Newman, 2013). Imaginemos un experimento que utiliza como estímulo un programa de televisión antisocial que ha sido visto por una tercera parte del grupo al que se le expondrá, mientras que al otro grupo se le expone a un programa prosocial que nadie ha visto. Condiciones agresivas, dolorosas, de cansancio, etcétera, pueden provocar mortalidad diferencial entre los grupos, la cual puede ocurrir no solo por el experimento en sí, sino por el tipo de personas que componen cada grupo o por factores externos al experimento. En experimentos médicos, esta posibilidad es una amenaza ante la cual el investigador debe estar muy pendiente.
10. *Interacción entre selección y maduración.* En este caso se trata de un efecto de maduración desigual en los grupos del experimento, debido a algún factor de selección, el cual puede dar origen a diferentes tasas de maduración o cambio autónomo entre grupos. Por ejemplo, si un grupo está compuesto por personas que acostumbran alimentarse a cierta hora (12:00 p.m.) y el otro, por personas que se alimentan a otra hora (3:00 p.m.), y el experimento se lleva a cabo de 11:00 a.m. a 2:30 p.m., la selección tendenciosa puede provocar un efecto de maduración distinto en los dos grupos:

hambre. Este sesgo se puede presentar si en una convención de una empresa multinacional experimentamos con ejecutivas de distintos países latinoamericanos (cuyos horarios de alimentación son muy distintos), y no tomamos en cuenta la interacción que puede producirse entre la selección y la maduración.

11. *Difusión de tratamientos experimentales.* En este caso, puede ser que los participantes de los grupos experimentales y de control intercambien información sobre la naturaleza del experimento, en particular respecto del estímulo, lo cual podría “nublar” los efectos de este. Por ejemplo, si ambos grupos saben que uno de ellos recibe dinero por participar y el otro no, los integrantes de este último podrían protestar y perturbar a toda la muestra, lo cual afectaría al experimento. Si en el caso de Naves y Poplawsky (1984), los estudiantes hubieran comentado a los demás participantes que se les “aparecía” un deficiente mental, los siguientes individuos no actuarían con naturalidad y el experimento resultaría un fracaso. Esta fuente resulta un riesgo cuando los grupos del experimento mantienen cercanía o contacto, pero se puede evitar si se eligen participantes que pertenezcan a diferentes ámbitos, pero equivalentes (en el ejemplo, estudiantes de diferentes carreras y semestres). Cuando se sospeche que los tratamientos se difunden, conviene entrevistar a los participantes para evaluar el grado de comunicación entre grupos y sus consecuencias.
12. *Actuaciones anormales del grupo de control.* Consiste en que si el grupo testigo conoce su condición y puede esforzarse en un grado superlativo por obtener puntuaciones más favorables en la variable dependiente. Por ejemplo, en un experimento para probar un nuevo método educativo, los participantes del grupo que no aprende por medio de este enfoque innovador, al deducir que se espera de ellos que obtengan las puntuaciones más bajas, se esfuerzan más de lo normal y obtienen calificaciones elevadas. O bien, pacientes del grupo de control que al saber que no reciben el estímulo, cuiden mucho más su alimentación, se automediquen, hagan más ejercicio, etcétera.
13. *Compensación.* Los beneficios de un experimento pueden ser desiguales o inequitativos para los grupos, lo que puede provocar que el grupo de control lo resienta y esto afecte los resultados (Creswell, 2013a). Por ejemplo, si el grupo experimental recibe una terapia o tratamiento médico y el de control, no, los miembros de este último pueden sentirse devaluados o desmoralizados, situación particularmente crítica cuando a un grupo se le otorgan estímulos económicos y al otro, no, como ya se señaló. Por ejemplo, en un experimento para analizar si un equipo de trabajo incrementa significativamente su productividad y calidad cuando se le asigna una nueva maquinaria, para lo cual se comparan sus resultados con los de otra cuadrilla que aún trabaja con el equipo antiguo, podría presentarse la compensación o una actuación anormal de cualquiera de los grupos.
14. *Otras interacciones.* Es posible que se presenten diversos efectos provocados por la interacción de las fuentes de invalidación interna. La selección puede interactuar con la mortalidad experimental, la historia con la maduración, la maduración con la inestabilidad, etc. También pueden operar simultáneamente varias de estas fuentes y la validez interna se deteriora aún más. Cada vez que hay mayor presencia de tales amenazas, crece la incertidumbre con respecto a las causas que produjeron cambios en las variables dependientes. Si no hubo cambios, no podemos saber si se debió a que una

o más fuentes contrarrestaron los posibles efectos del tratamiento. Por ejemplo, un método de enseñanza exitoso se utilizó con los participantes menos motivados y el menos exitoso con los más motivados, lo cual provocó que se compensaran ambos factores. De este modo, no sabremos cómo interpretar los resultados.

Los sujetos participantes y el experimentador como fuentes de invalidación interna

Otra razón que atenta contra la interpretación correcta y certera de los resultados de un experimento en el que se involucran personas, es el comportamiento de estas. Es posible que ellas ingresen al experimento con ciertas actitudes, expectativas y prejuicios que alteren su comportamiento durante el estudio, por ejemplo, no colaborar y actuar negativamente, hasta el punto de llegar a ser hostiles. Este posible comportamiento debe tenerse presente antes y durante la investigación. Es necesario evaluar qué participantes pueden arruinar el experimento y descartarlos, o procurar que en todos los grupos haya personas con actitudes positivas y negativas (si todos los que tienen actitudes negativas integran un grupo, la validez interna estaría en problemas). Recordemos que los individuos que intervienen en un experimento, de una manera u otra, tienen motivos para participar y su conducta muchas veces puede afectar la calidad de las interpretaciones.

Además, el mismo experimentador no es un observador pasivo que no interactúa, sino un sujeto activo que puede influir en los resultados del estudio (Christensen, 2006). El investigador tiene una serie de motivos que lo impulsan a realizar su experimento, pues con los resultados que este genere puede probar sus hipótesis. Ello, conscientemente o no, puede conducirlo a afectar el comportamiento de los sujetos en dirección de su hipótesis. Por ejemplo, dar explicaciones más completas a uno de los grupos. Hay que evitar todas estas posibles desviaciones. Por ello, en ciertos casos quien trate con los participantes no debe ser el experimentador, sino alguien que no conozca la hipótesis, las condiciones experimentales ni los propósitos del estudio, sino que solo reciba instrucciones precisas sobre lo que debe hacer y cómo desempeñarse. En ciencias exactas, a pesar de que se experimenta con sustancias o fuerzas, el investigador que manipula las condiciones también debe estar consciente de su rol.

Quienes participan en el experimento tampoco deben conocer las hipótesis ni las condiciones experimentales; incluso, con frecuencia es necesario distraerlos de los verdaderos propósitos del estudio, aunque al finalizar se les deba dar una explicación completa de la investigación. Cuando, por ejemplo, se analizan los efectos de ciertos medicamentos, los investigadores hacen creer a un grupo que se le está administrando una medicina cuando en realidad no es así, sino que se le proporciona píldoras de azúcar. Este “truco” evita la influencia que la expectativa de recibir medicamento pudiera tener en la variable dependiente. A esta sustancia que no tiene efectos se le denomina “placebo”.

Con métodos de instrucción, por ejemplo, ocurre que el grupo que se expondrá al método innovador es influido por el simple hecho de decirle que se trata de un nuevo método. Lo mismo con pruebas de sabor de un producto alimenticio, programas de televisión, experiencias motivacionales, diseños arquitectónicos, sistema de calidad en las empresas, etc. Por tanto, estos aspectos deben tomarse muy en cuenta.

Varios grupos de comparación

En el capítulo siete del texto impreso y en líneas anteriores, se hace hincapié en que para lograr la validez interna y el control en un experimento es necesario que se aplique, por lo menos, a dos grupos comparables. En primer término, porque si se trabaja con un solo grupo no es posible saber con certeza si influyeron las fuentes de invalidación interna o no. Por ejemplo, mediante un experimento intentamos probar la hipótesis que sostiene que “a mayor información psicológica sobre una clase social, menor prejuicio hacia ella”. Si decidimos trabajar con un solo grupo, se debería exponerlo a un programa de sensibilización donde se proporcione información sobre la manera en que vive dicha clase, sus angustias y problemas, necesidades, sentimientos, aportaciones a la sociedad, etc., para luego observar el nivel de prejuicio (el programa debe incluir charlas de expertos, películas y testimonios grabados, lecturas, etc.). Este experimento se esquematizaría así:

	Momento 1 (X)	Momento 2 (O)
G	Exposición al programa de sensibilización	Observación del nivel de prejuicio

Todo en un único grupo. ¿Qué sucede si se observa un bajo nivel de prejuicio en el grupo? ¿Podemos deducir con absoluta certeza que se debió al estímulo? Desde luego que no. Es posible que ese bajo nivel se deba al programa de sensibilización, que es la forma de manipular la variable independiente “información psicológica sobre una clase social”, pero también a que los participantes tenían un bajo nivel de prejuicio antes del experimento y, en realidad, el programa no afectó. Además, no lo podemos saber porque no hay una medición del nivel de prejuicio al inicio del experimento (antes de la presentación del estímulo experimental); es decir, no existe punto de comparación. Pero, aunque hubiera ese punto de contraste inicial, con un solo grupo no podríamos estar seguros de cuál fue la causa del nivel de prejuicio. Supongamos que este nivel de prejuicio antes del estímulo o tratamiento era alto, y después del estímulo, bajo. Quizás el tratamiento sea la causa del cambio, pero tal vez también ocurrió lo siguiente:

1. Que la primera prueba de prejuicio sensibilizara a los sujetos participantes y que influyera en sus respuestas en la segunda prueba. De este modo, las personas crearon conciencia de lo negativo de ser prejuiciosas cuando respondieron a la medición previa (administración de prueba).
2. Que los individuos seleccionados se agotaran durante el experimento y sus respuestas a la segunda prueba fueran “a la ligera” (maduración).
3. Que los sujetos prejuiciosos o parte importante de ellos hubiera abandonado el experimento (mortalidad experimental).

También podrían existir otras razones. Además, si no se hubiera observado un cambio en el nivel de prejuicio entre la primera prueba (antes del programa) y la segunda (después del programa), se podría asumir que la exposición al programa no tiene efectos, aunque también podría suceder que el grupo seleccionado es muy prejuicioso y tal vez el programa sí tiene efectos en personas con niveles comunes de prejuicio. Asimismo, si el cambio es negativo (mayor nivel de prejuicio en la segunda medición que en la primera), se podría

suponer que el programa incrementa el prejuicio, pero supongamos que haya ocurrido un suceso durante el experimento que generó momentáneamente prejuicios hacia esa clase social (una violación en la localidad a cargo de un individuo de esta), pero después los participantes “regresaron” a su nivel de prejuicio normal (regresión). Incluso podría haber otras explicaciones.

Con un solo grupo no estaríamos seguros de que los resultados se debieran al estímulo experimental o a otras razones. Siempre quedará la duda. Los “experimentos” con un grupo se basan en sospechas o en lo que “aparentemente es”, pero carecen de fundamentos científicos. Cuando se trabaja con un solo grupo se corre el riesgo de seleccionar sujetos atípicos (los más inteligentes al experimentar con métodos de enseñanza, los trabajadores más motivados cuando se aplican programas de incentivos, los consumidores más críticos, las parejas de novios más integradas, etc.) y de que intervengan la historia, la maduración y demás fuentes de invalidación interna, sin que el experimentador se percate de ello.

Por todas estas razones, el investigador debe tener al menos un punto de comparación: dos grupos, uno al que se le administra el estímulo y otro que no lo reciba (el grupo de control). Tal como se mencionó al hablar de manipulación, a veces se debe trabajar con diversos grupos cuando se desea averiguar el efecto de distintos niveles o modalidades de la variable independiente.

Equivalencia de los grupos en experimentos

Para tener control no basta con dos o más grupos, sino que estos deben ser similares en todo, menos en la manipulación de la o las variables independientes. El control implica que todo permanece constante, salvo tal intervención. Si entre los grupos que conforman el experimento todo es similar o equivalente, excepto la manipulación de la variable independiente, las diferencias entre ellos pueden atribuirse a esta y no a otros factores (entre los cuales están las fuentes de invalidación interna).

Imaginemos que deseamos indagar si una serie de programas educativos de televisión para niños genera mayor aprendizaje que un método tradicional. Un grupo recibe la enseñanza a través de los programas, otro la recibe por medio de instrucción oral tradicional, mientras que un tercero dedica ese mismo tiempo a jugar libremente en el salón de clases. Supongamos que los niños que aprendieron mediante los programas obtienen las mejores calificaciones en una prueba de conocimientos relativa a los contenidos enseñados, los que recibieron el método tradicional obtienen calificaciones mucho más bajas, y los que jugaron obtienen puntuaciones de cero o cercanas a este valor. En forma aparente, los programas son un mejor vehículo de enseñanza que la instrucción oral. Pero si los grupos no son equivalentes, no podemos confiar en que las diferencias se deban realmente a la manipulación de la variable independiente (programas televisivos-instrucción oral) y no a otras causas, o a la combinación de ambas. Por ejemplo, a los niños más inteligentes, estudiosos y con mayor empeño se les asignó al grupo que fue instruido por televisión, o simplemente su promedio de inteligencia y aprovechamiento era el más elevado; o la instructora del método tradicional no poseía buen desempeño, o los niños expuestos a este último método recibieron mayor carga de trabajo y tenían exámenes los días en que se

desarrolló el experimento, etc. ¿Cuánto se debió al método y cuánto a otros factores? Para el investigador la respuesta a esta pregunta se convierte en un enigma: no hay control.

Si experimentáramos con métodos de motivación para trabajadores, y un grupo estuviera conformado por los que laboran en el turno matutino, y el otro por los empleados del turno vespertino, ¿quién nos asegura que antes de iniciar el experimento ambos tipos de trabajadores están igualmente motivados? Puede haber discrepancias en la motivación inicial porque los supervisores de distintos turnos motivan de diferente manera y grado, o tal vez los del turno vespertino preferirían trabajar en la mañana o se les pagan menos horas extras, etc. Si no están igualmente motivados, podría ocurrir que el estímulo aplicado a los del turno de la mañana aparentara ser el más eficaz, cuando en realidad no es así.

Veamos un ejemplo que nos ilustrará el efecto tan negativo que puede generar la inequivalencia de los grupos sobre los resultados de un experimento. ¿Qué investigador probaría el efecto de diferentes métodos para sensibilizar a las personas respecto de lo terrible que puede ser el terrorismo si un grupo está constituido por miembros de Al-Qaeda y el otro por familiares de las víctimas de los atentados en Londres, en julio de 2005?

Los grupos deben ser equivalentes cuando se inicia y durante todo el desarrollo del experimento, menos en lo que respecta a la variable independiente. Asimismo, los instrumentos de medición deben ser iguales y aplicados de la misma manera.

Equivalencia inicial

Este requisito implica que los grupos son similares entre sí en el momento de iniciarse el experimento. Si este versa sobre los métodos educativos, los grupos deben ser equiparables en cuanto a número de personas, inteligencia, aprovechamiento, disciplina, memoria, género, edad, nivel socioeconómico, motivación, nutrición, conocimientos previos, estado de salud física y mental, interés por los contenidos, extraversión, etc. Si inicialmente no son equiparables, digamos en cuanto a motivación o conocimientos previos, las diferencias entre los grupos —en cualquier variable dependiente— no podrían atribuirse con certeza a la manipulación de la variable independiente.

La equivalencia inicial no se refiere a equivalencias individuales, porque todos diferimos de manera natural en muchos aspectos, sino a la equivalencia entre grupos. Si trabajamos con dos grupos, es indudable que habrá, por ejemplo, personas muy inteligentes en uno de ellos, pero también debe haberlas en el otro. Si en uno hay mujeres, en el otro debe haberlas en la misma proporción. Y así con todas las variables que puedan afectar a la o las variables dependientes, además de la variable independiente. Para hablar de experimentos válidos no pueden existir diferencias significativas entre los grupos.

Equivalencia durante el experimento

Durante el estudio, los grupos deben mantener la equivalencia inicial en los aspectos concernientes al desarrollo experimental, excepto en la manipulación de la variable independiente: mismas instrucciones (salvo variaciones que sean parte de esa manipulación), personas con las que tratan los participantes y maneras de recibirlos, lugares con características semejantes (iguales objetos en los cuartos, clima, ventilación, sonido ambiental, etcétera),

Equivalencia inicial Implica que los grupos son similares entre sí en el momento de iniciarse el experimento.





misma duración del experimento, así como del momento y, en fin, todo lo que sea parte del estudio. Cuanto mayor sea la equivalencia, durante su desarrollo, habrá mayor control y posibilidad de que, si observamos o no efectos, estemos seguros de que verdaderamente los hubo o no.

Cuando trabajamos simultáneamente con varios grupos, es difícil que las personas que dan las instrucciones y vigilan el desarrollo de los grupos sean las mismas. En este caso debe buscarse que su tono de voz, apariencia, edad, género y otras características capaces de afectar los resultados sean iguales o similares, y mediante entrenamiento debe estandarizarse su proceder. Algunas veces se dispone de menos cuartos o lugares que de grupos. En consecuencia, la asignación de los grupos a los cuartos y horarios se realiza al azar, pero se debe procurar que los procedimientos se apliquen, dentro de lo posible, en horarios coincidentes. Otras veces, los participantes reciben los estímulos individualmente y su exposición no puede ser simultánea. Se deben sortear de manera que en un día (por la mañana) personas de todos los grupos participen en el experimento, lo mismo por la tarde y durante el tiempo que sea necesario (los días que dure el experimento).

Recordatorio: ¿cómo se logra la equivalencia inicial?

Asignación al azar

Como se comentó en el capítulo 7 de la obra impresa, el método más exacto para lograr la equivalencia inicial es la asignación aleatoria o al azar de los participantes a los grupos del experimento (en inglés, randomization). La asignación al azar nos asegura probabilísticamente que dos o más grupos son equivalentes entre sí (Kirk, 2012; Knapp, 2008; Pettygrove, 2007 y Peng, 2003). Esta técnica de control tiene como propósito dar al investigador la seguridad de que variables extrañas, conocidas o desconocidas, no afectarán de manera sistemática los resultados del estudio (Christensen, 2006). Esta técnica, que fue diseñada por Sir Ronald A. Fisher, en la década de 1940, ha demostrado durante años que funciona para hacer equivalentes a grupos de participantes. Como mencionan Cochran y Cox (1992), la asignación aleatoria es en cierta forma análoga a un seguro, por el hecho de que es una precaución contra interferencias que pueden o no ocurrir, y ser o no importante si suceden. Por lo general, es aconsejable tomarse el trabajo de distribuir aleatoriamente, aun cuando no se espere que haya un sesgo importante si se opta por otro enfoque.

Si la única diferencia que distingue al grupo experimental y al de control es la variable independiente, las diferencias entre ellos pueden atribuirse a esta última. Pero si hay otras diferencias, no podríamos hacer tal afirmación.

Ejemplo de equivalencia inicial de los grupos en experimentos

En el capítulo 7 del libro impreso se señaló que en un experimento, los grupos de comparación deben poseer equivalencia inicial.¹ A continuación ejemplificamos esta noción.

¹ En experimentos en ciencias biológicas, obviamente los ejemplares de los grupos que se contrastan también deben ser similares, y en experimentos con materiales, sustancias, fuerzas y objetos en ciencias físicas, por ejemplo, ocurre lo mismo (incluso idénticos, por ello les denominamos “ciencias exactas”). La equivalencia inicial es una condición de la experimentación.

Si suponemos que todas las variables (V) pudieran medirse de uno a 10, la equivalencia entre grupos se debería conceptualizar tal como en la **tabla 5.1**.

● **TABLA 5.1** Esquema para ilustrar la equivalencia inicial de los grupos.

Al inicio del experimento	
Grupo 1	Grupo 2
V1 = 8	V1 = 7.9
V2 = 4	V2 = 4.1
V3 = 6	V3 = 6
V4 = 7.2	V4 = 7.4
V5 = 10	V5 = 9.9
20 mujeres, 21 hombres	19 mujeres, 22 hombres
Promedio de edad = 25 años, 6 meses	Promedio de edad = 25 años, 4 meses
V6 = 2	V6 = 2
Vk = K	Vk = K
V = una variable (V1 = variable 1, V2 = variable 2...)	

Veamos en la **tabla 5.2** un ejemplo de equivalencia entre dos grupos respecto de ciertos rasgos físicos para ilustrar el concepto.

● **TABLA 5.2** Ejemplo de dos grupos equivalentes.

Grupo 1		Grupo 2
10 hombres de tez morena y ojos cafés	equivalencia =	10 hombres de tez morena y ojos cafés
5 hombres de tez morena y ojos negros		4 hombres de tez morena y ojos negros
11 hombres de tez clara y ojos cafés		12 hombres de tez clara y ojos cafés
5 hombres de tez clara y ojos verdes		5 hombres de tez clara y ojos verdes
10 mujeres de pelo rubio y tez muy clara		10 mujeres de pelo rubio y tez muy clara
8 mujeres de pelo castaño claro y tez clara		9 mujeres de pelo castaño claro y tez clara
5 hombres de pelo castaño oscuro y tez clara		3 hombres de pelo castaño oscuro y tez clara

En la **tabla 5.3** se presenta un ejemplo de dos grupos que interculturalmente no son equivalentes.

● **TABLA 5.3** Ejemplo de dos grupos no equivalentes.

Grupo 1		Grupo 2
3 venezolanos	≠	1 venezolano
6 colombianos		3 brasileñas
5 mexicanos		2 mexicanos
6 estadounidenses		1 estadounidense
4 ingleses		28 franceses
7 bolivianos		10 ingleses
3 italianos		4 rusos
5 israelitas		2 alemanes
4 afganos		5 suizos
3 cubanos		2 nicaragüenses
12 españoles		



Desde luego, es prácticamente imposible alcanzar la equivalencia perfecta o ideal con seres humanos o animales, pero no deben permitirse diferencias iniciales significativas entre los grupos.

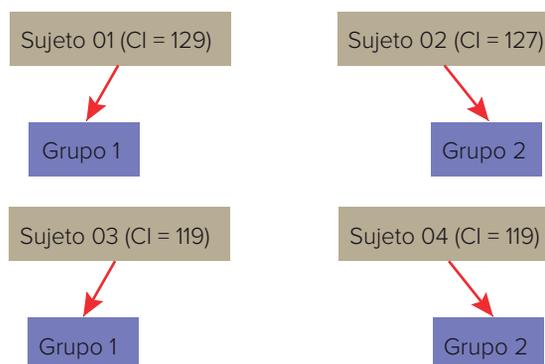
Ejemplo de la técnica de emparejamiento en experimentos

A fin de explicar aún más lo que se expuso en el capítulo 7 del libro sobre el emparejamiento, a continuación presentamos un ejemplo con dos grupos.

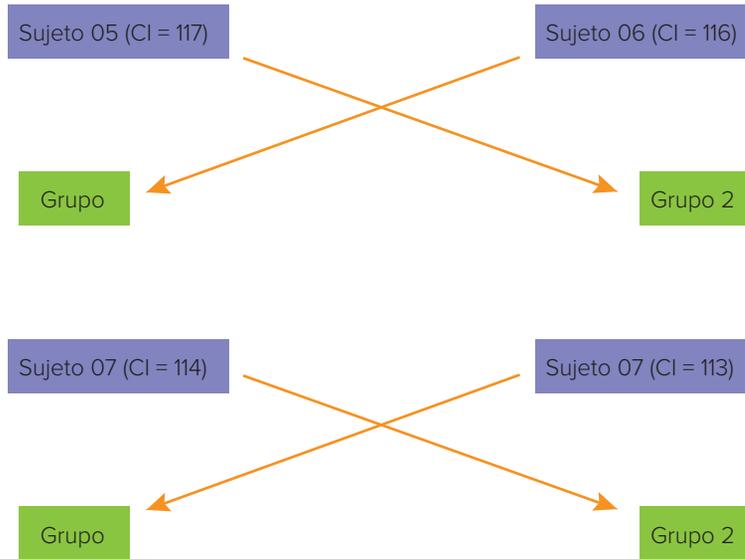
Suponga que se tienen 16 personas para un experimento sobre desempeño académico (recuerde la sugerencia de tener 15 o más en cada grupo; aquí se incluyen 16 solo para abreviar el caso), divididas en un grupo al cual se le otorgará una beca de 50% para sus estudios universitarios, y un grupo de control al que no se le concederá beneficio económico alguno. La variable dependiente sería el desempeño académico, medido por el promedio de sus calificaciones en todas las asignaturas. Los estudiantes se ordenarían de la siguiente manera:

Participante	Coficiente de inteligencia (CI)	Participante	Coficiente de inteligencia (CI)
01	129	09	110
02	127	10	110
03	119	11	108
04	119	12	107
05	117	13	106
06	116	14	105
07	114	15	104
08	113	16	102

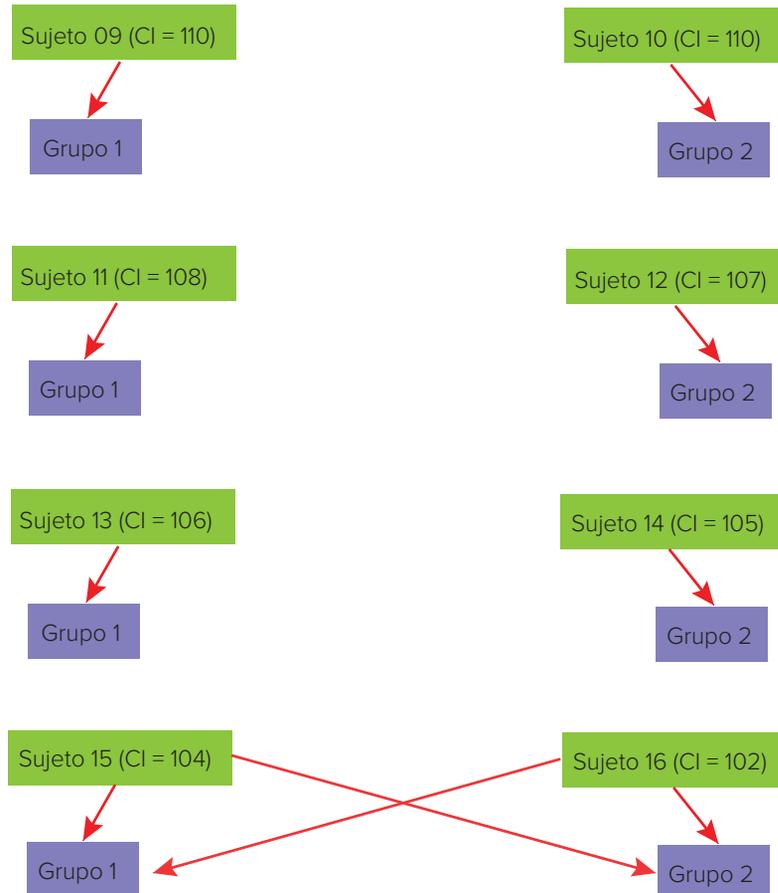
Se procede entonces a emparar gradualmente a dos grupos en la variable deseada, en este caso, inteligencia:



Hasta ahora, el grupo 1 lleva dos puntos más que el grupo 2 (grupo 1 = 248, grupo 2 = 246), por lo cual es necesario compensarlo.



Hasta aquí se ha conservado el balance entre los grupos (grupo 1 = 477 y grupo 2 = 477).





De esta forma, los grupos quedarían emparejados en inteligencia.

Grupo 1		Grupo 2	
Sujeto	CI	Sujeto	CI
S01	129	S02	127
S03	119	S04	119
S06	116	S05	117
S08	113	S07	114
S09	110	S10	110
S11	108	S12	107
S13	106	S14	105
S16	102	S15	104
Promedio = 112.87		Promedio = 112.87	

Ambos grupos son equivalentes en cuanto a la variable deseada. Este procedimiento puede extenderse a más de dos grupos (si se trata de tres, en lugar de parejas se iguala por tercias, con cuatro grupos por cuartetos, etcétera).

Respecto del emparejamiento debemos recordar lo que señalamos en el texto impreso: la asignación al azar es el mejor método para hacer equivalentes los grupos (más preciso y confiable). El emparejamiento no la sustituye por completo, pues llega a suprimir o eliminar el posible efecto de la variable apareada, pero nunca nos asegura que otras variables (no apareadas) no puedan afectar los resultados del experimento. En cambio, la asignación aleatoria garantiza que otras variables (además de la o las variables independientes de interés para el investigador) no afectan las dependientes ni confundan al experimentador. Como comentan The SAGE glossary of the social and behavioral sciences (2009s), Knapp (2008) y Nunnally y Bernstein (1994), lo bueno de la asignación al azar de los participantes o casos a los grupos de un diseño experimental es que el procedimiento asegura absolutamente que, en promedio, los grupos no diferirán (antes de que participen en los tratamientos experimentales) en ninguna característica más de lo que pudiera esperarse por pura casualidad.

Diseño de cuatro grupos de Solomon: ampliación

Para iniciar, recordemos al lector que este diseño es una mezcla de dos diseños: 1) diseño con posprueba únicamente y grupo de control y 2) diseño de preprueba-posprueba con grupo de control. La suma de estos dos diseños origina cuatro grupos: dos experimentales y dos de control, pero mientras los primeros reciben el mismo tratamiento experimental, los segundos no reciben ningún tratamiento. Solo a uno de los grupos experimentales y a uno de los grupos de control se les administra la preprueba; a los cuatro grupos se les aplica la posprueba. Todos los participantes son asignados al azar.

Tengamos en mente que el diseño se diagrama así:

RG ₁	O ₁	X	O ₂
RG ₂	O ₃	—	O ₄
RG ₃	—	X	O ₅
RG ₄	—	—	O ₆

El diseño original incluye solo cuatro grupos y un tratamiento experimental. Los efectos se determinan comparando las cuatro pospruebas. Los grupos uno y tres son experimentales, y los grupos dos y cuatro son de control.

La ventaja de este diseño es que el experimentador tiene la posibilidad de verificar los posibles efectos de la preprueba sobre la posprueba, puesto que a unos grupos se les administra un test previo y a otros no. Es posible que la preprueba afecte la posprueba o que aquélla interactúe con el tratamiento experimental. Por ejemplo, con promedios de una variable determinada podría encontrarse lo que muestra la **tabla 5.4**.

● **TABLA 5.4** Ejemplo de efecto de preprueba en el diseño de Solomon.

RG ₁	O ₁ = 8.0	X	O ₂ = 14.0
RG ₂	O ₃ = 8.1	—	O ₄ = 11.0
RG ₃	—	X	O ₅ = 11.0
RG ₄	—	—	O ₆ = 8.0

Teóricamente, O_2 debería ser igual a O_5 , porque ambos grupos recibieron el mismo tratamiento; asimismo, O_4 y O_6 deberían tener el mismo valor, porque ninguno recibió estímulo experimental. Pero $O_2 \neq O_5$ y $O_4 \neq O_6$. Entonces, ¿cuál es la única diferencia entre O_2 y O_5 y entre O_4 y O_6 ? La respuesta es la preprueba. Las diferencias pueden atribuirse a un efecto de la preprueba (la preprueba impacta, aproximadamente, tres puntos, y el tratamiento experimental también tres puntos, poco más o menos). Veámoslo de manera esquemática:

Ganancia con preprueba y tratamiento = 6

Ganancia con preprueba y sin tratamiento = 2.9 (casi 3).

Debido a que, al inicio, la técnica de distribución aleatoria hace equivalentes a los grupos, supuestamente el promedio de la preprueba hubiera sido para todos cerca de ocho, si se hubiera aplicado a los cuatro grupos. La “supuesta ganancia” (supuesta porque no hubo preprueba) del tercer grupo, con tratamiento y sin preprueba, es de tres. Y la “supuesta ganancia” (supuesta porque tampoco hubo preprueba) del cuarto grupo es nula o inexistente (cero).



Estas conclusiones indican que cuando hay preprueba y estímulo se obtiene la máxima puntuación de 14, si solo hay preprueba o estímulo la puntuación es de 11, y cuando no hay preprueba ni estímulo es de ocho (calificación que todos deben tener inicialmente por efecto de la asignación al azar). También podría ocurrir un resultado como el de la **tabla 5.5**.

● **TABLA 5.5** Ejemplo del efecto de interacción entre la preprueba y el estímulo en el diseño de Solomon.

RG ₁	O ₁ = 7.9	X	O ₂ = 14.0
RG ₂	O ₃ = 8.0	—	O ₄ = 8.1
RG ₃	—	X	O ₅ = 11.0
RG ₄	—	—	O ₆ = 7.9

En este caso, la preprueba no afecta (vea la comparación entre O₃ y O₄), pero el estímulo sí lo hace (compare O₅ con O₆); sin embargo, cuando el estímulo o tratamiento se junta con la preprueba se observa un efecto importante (compare O₁ con O₂), un efecto de interacción entre el tratamiento y la preprueba.

El diseño de Solomon controla todas las fuentes de invalidación interna por las mismas razones que fueron explicadas en diseños “puros” del texto impreso. La administración de pruebas se somete a un análisis minucioso.

Fuentes de invalidación externa a detalle

En el texto impreso se dijo que se ampliarían las fuentes de invalidación externa en el centro de recursos en línea, por lo que aquí se hace.

Efecto reactivo o de interacción de las pruebas

Este efecto se presenta cuando la preprueba aumenta o disminuye la sensibilidad o la calidad de la reacción de los participantes a la variable experimental, lo cual contribuye a que los resultados obtenidos para una población con preprueba no puedan generalizarse a quienes forman parte de esa población sin preprueba. Babbie (2017) utiliza un excelente ejemplo de esta influencia: en un experimento diseñado para analizar si una película disminuye el prejuicio racial, la preprueba podría sensibilizar al grupo experimental y la película podría lograr un efecto mayor del que tendría si no se aplicara la preprueba (por ejemplo, si se pasara la película en un cine o en la televisión).

Efecto de interacción entre los errores de selección y el tratamiento experimental

Este factor se refiere a que se elijan personas con una o varias características que permitan que el tratamiento experimental produzca un efecto, el cual no se lograría si las personas no tuvieran esas peculiaridades. Por ejemplo, si seleccionamos trabajadores muy motivados para un experimento sobre productividad, podría ocurrir que el tratamiento solo tuviera efecto en este tipo de trabajadores y no en otros (solo funciona con individuos sumamente motivados). Ello se resolvería con una muestra representativa de todos los trabajadores o

mediante la introducción de un diseño factorial y que una de las variables fuera el grado de motivación (vea más adelante diseños factoriales).

En ocasiones, este factor se presenta cuando se reclutan voluntarios para la realización de algunos experimentos.

Efectos reactivos de los tratamientos experimentales

La “artificialidad” de las condiciones puede hacer que el contexto experimental resulte atípico respecto de la manera en que se aplica regularmente el tratamiento (Campbell, 1975). Por ejemplo, a causa de la presencia de observadores y equipo, los participantes cambian su conducta normal en la variable dependiente medida, la cual no se alteraría en una situación común donde se aplicara el tratamiento. Por ello, el experimentador tiene que ingeniárselas para hacer que los sujetos se olviden de que están en un experimento y no se sientan observados. A esta fuente también se le conoce como “efecto Hawthorne”, por una serie de experimentos muy famosos desarrollados entre 1924 y 1927, en una planta del mismo nombre de la Western Electric Company, en la cual, al variarse la intensidad de la iluminación se observaban incrementos de la productividad de los trabajadores, tanto al aumentarla como al disminuirla. Por lo tanto, se concluyó que los cambios en la productividad se debían a que los participantes se sentían observados (Ballantyne, 2000).

Interferencia de tratamientos múltiples

Si se aplican varios tratamientos a un grupo experimental para conocer sus efectos por separado y en conjunto (por ejemplo, enseñarle a niños hábitos higiénicos con un video, más una dinámica que implique juegos, más un libro explicativo); incluso, si los tratamientos no son de impacto reversible, es decir, si no es posible eliminar sus efectos, las conclusiones solo podrán hacerse extensivas a los infantes que experimenten la misma secuencia de tratamientos, sean múltiples o repetitivos (vea en páginas subsecuentes los diseños con diversos tratamientos).

Imposibilidad de replicar los tratamientos

Cuando los tratamientos son tan complejos que no pueden replicarse en situaciones no experimentales, es difícil generalizar sus efectos.

Descripciones insuficientes del tratamiento experimental

En ocasiones, el tratamiento o los tratamientos experimentales no se describen con suficiente detalle en el reporte del estudio, por lo que si otro investigador desea reproducirlos, le resultará muy difícil o imposible hacerlo (Mertens, 2015). Por ejemplo, señalamientos tales como “la intervención funcionó” no nos dice nada, por lo cual se debe especificar en qué consistió tal intervención. Las instrucciones deben incluirse, y la minuciosidad con que se comuniquen es un elemento importante.

Efectos de novedad e interrupción

Un nuevo tratamiento puede tener resultados positivos simplemente por ser percibido como novedoso, o bien, lo contrario: tener un efecto negativo porque interrumpe las actividades normales de los participantes. En este caso, es recomendable inducir a los



sujetos paulatinamente al tratamiento (no de manera intempestiva) y esperar a que asimilen los cambios que este naturalmente provoca (Gerber y Green, 2012 y Mertens, 2015).

El experimentador

A quien también consideramos una fuente de invalidación interna, pues puede generar alteraciones o cambios que no se presentan en situaciones no experimentales. Es decir, que el tratamiento solo tenga efecto con la intervención del experimentador.

Interacción entre la historia o el lugar y los efectos del tratamiento experimental

Un experimento conducido en un contexto en particular (tiempo y lugar), en ocasiones no puede ser duplicado (Mertens, 2015). Por ejemplo, un estudio que se efectúe en una empresa en el momento en que se reestructuran departamentos (donde algunos quizá se mantengan, otros se reduzcan y hasta algunos desaparezcan). O bien, un experimento en una escuela secundaria que se lleva a cabo al tiempo que su equipo de fútbol obtiene un campeonato nacional.

Asimismo, en ocasiones los resultados de un experimento no pueden generalizarse a otros lugares o ambientes. Si se realiza una investigación en una escuela pública recientemente inaugurada, que cuenta con los máximos avances tecnológicos y educativos, ¿podemos extrapolar los resultados a todas las escuelas públicas de la localidad? A veces, el efecto del tratamiento lo tenemos que analizar en distintos lugares y tiempos (Creswell, 2013a).

Mediciones de la variable dependiente

Puede suceder que un instrumento no registre cambios en la variable dependiente (ejemplo: cuestionario) y otro, sí (observación). Si en un experimento se utiliza un instrumento para recolectar datos, y de este modo sus resultados puedan compararse, otros estudios deberán evaluar la variable dependiente con el mismo instrumento o uno equivalente (lo mismo en situaciones no experimentales).

Diseños experimentales de series cronológicas múltiples

Cuando el experimentador se interesa en analizar los *efectos en el mediano o largo plazos*, porque supone que la influencia de la variable independiente sobre la dependiente tarda en manifestarse, por ejemplo, en programas de difusión de innovaciones, tratamientos médicos, métodos educativos, procesos industriales, introducción de tecnologías o estrategias psicoterapéuticas, *es conveniente que adopte diseños con varias pospruebas*. A estos

diseños se les conoce como *series cronológicas experimentales*.

En realidad, el término **serie cronológica** se aplica a cualquier diseño que permita efectuar varias observaciones o mediciones sobre una variable a través del tiempo, sea o no experimental, solo que en este caso se les llama experimentales porque reúnen los requisitos para serlo.

Estos diseños operan con dos o más grupos y los participantes o casos son asignados al azar a ellos. Sin embargo, debido a que transcurre mucho más tiempo entre el inicio



Series cronológicas Diseños que, a través del tiempo, efectúan diversas observaciones o mediciones sobre una variable.

y la terminación del experimento, el investigador debe tener cuidado de que no ocurra algo que afecte de manera distinta a los grupos (con excepción de la manipulación de la variable independiente). Lo mismo sucede cuando el estímulo se aplica durante mucho tiempo (por ejemplo, programas motivacionales para trabajadores que duran semanas), pues con el paso del tiempo es más difícil mantener la equivalencia inicial de los grupos.

Un ejemplo de cómo se diagraman las series cronológicas experimentales se muestra en la **figura 5.1**.

Serie cronológica sin preprueba, con varias pospruebas y grupo de control

RG ₁	X ₁	O ₁	O ₂	O ₃
RG ₂	X ₂	O ₄	O ₅	O ₆
RG ₃	X ₃	O ₇	O ₈	O ₉
RG ₄	—	O ₁₀	O ₁₁	O ₁₂

Serie cronológica con preprueba, con varias pospruebas y grupo de control

RG ₁	O ₁	X ₁	O ₂	O ₀₃	O ₀₄
RG ₂	O ₅	X ₂	O ₆	O ₀₇	O ₀₈
RG ₃	O ₉	—	O ₁₀	O ₁₁	O ₁₂

Serie cronológica basada en el diseño de cuatro grupos de Solomon

RG ₁	O ₁	X	O ₂	O ₃
RG ₂	O ₄	—	O ₅	O ₆
RG ₃	—	X	O ₇	O ₈
RG ₄	—	—	O ₉	O ₁₀

● **Figura 5.1** Ejemplos de experimentos de series cronológicas.

Las pospruebas pueden ser tantas como se requiera y sea posible aplicar.

Asimismo, en otras ocasiones se desea analizar la evolución de los grupos antes y después del tratamiento experimental. En esta situación pueden incluirse varias prepruebas y pospruebas, en cuyo caso se tendrían esquemas como el siguiente:

R	G ₁	O ₁	O ₂	O ₃	X ₁	0	0	0
R	G ₂	O ₄	O ₅	O ₆	X ₂	0	0	0
R	G ₃	O ₇	O ₈	O ₉	—	0	0	0

Ejemplo

Diseño experimental de serie cronológica

Un consultor en cuestiones organizacionales estudia el efecto sobre la dedicación al trabajo que tiene la difusión de una serie de valores, que la directiva de una corporación considera deben ser implantados en la empresa.

Pregunta de investigación: "Cuanto más conozca el personal de una corporación los valores de esta (definidos por su directiva), ¿mostrará mayor dedicación al trabajo?"

Hipótesis de investigación: "El mayor conocimiento de los valores corporativos genera mayor dedicación al trabajo".



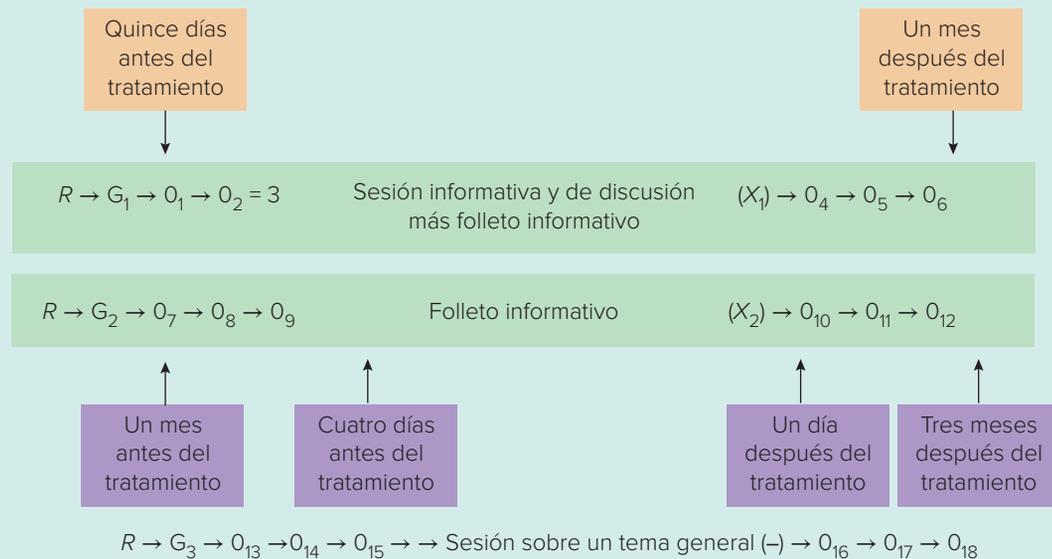
Hipótesis estadística: $rx_y > 0$

El consultor selecciona a 99 personas de la corporación, de todos los niveles jerárquicos, y los asigna al azar a tres grupos:

1. Uno participa en una reunión de dos horas en donde se le presentan ampliamente los valores corporativos con aplicaciones a situaciones específicas de trabajo; posteriormente sus miembros reciben un folleto con explicaciones adicionales.
2. Otro asiste a una sesión breve donde se proporciona un folleto sobre los valores corporativos (no hay explicación verbal ni discusión o sesión de preguntas y respuestas).
3. El tercer grupo asiste a una sesión donde se trata algún aspecto no relacionado con el trabajo o la organización (digamos, un tema de interés general).

Antes de administrar los tratamientos a todos los participantes, se les aplican tres mediciones de la dedicación al trabajo. Asimismo, después de los tratamientos, también se les aplica tres mediciones de la misma variable (al corto, mediano y largo plazos). El diseño se debería diagramar como se muestra en la **figura 5.2**.

Recuerde que se pretende medir la dedicación al trabajo. Cada grupo estaría integrado por 33 personas.

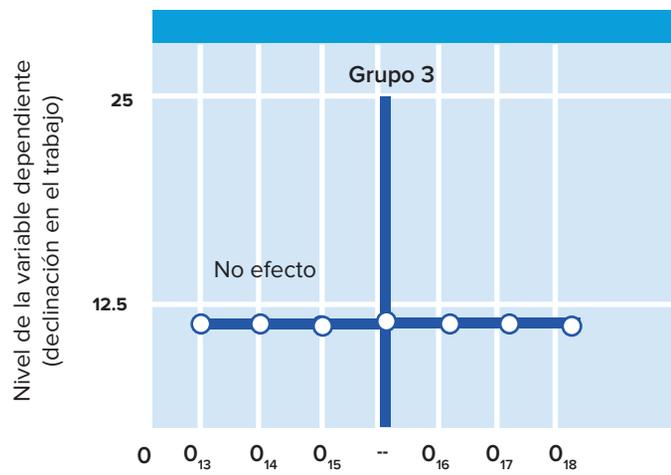
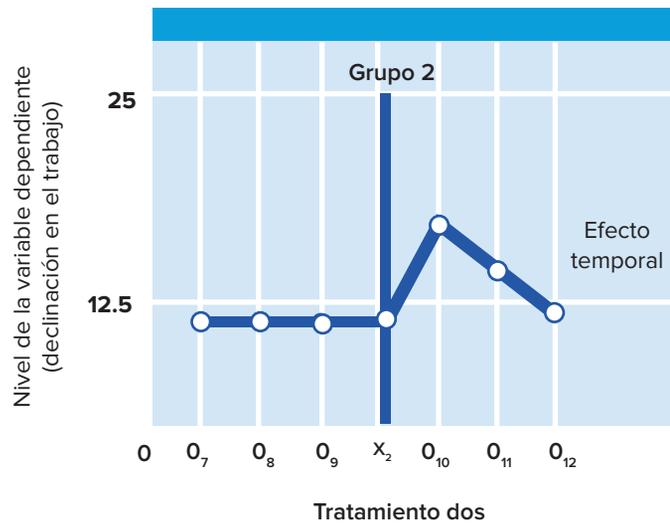
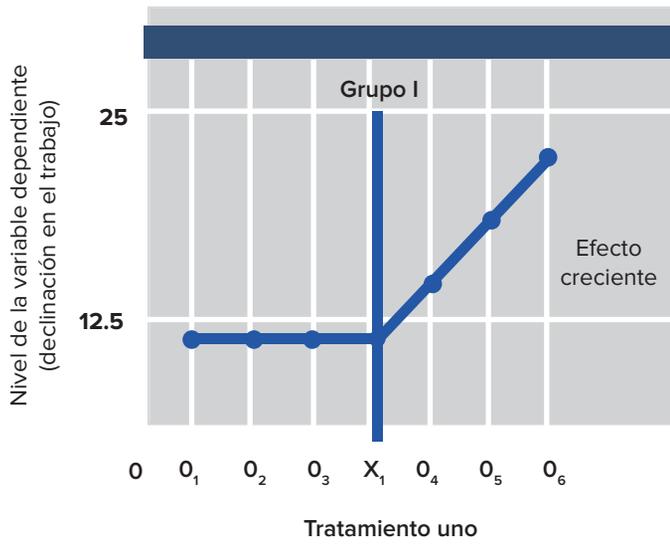


● **Figura 5.2** Diseño del ejemplo de serie cronológica sobre la dedicación al trabajo de tres grupos.

Una ventaja del diseño consiste en que es posible evaluar la evolución comparativa de los grupos. Por ejemplo, si se encontraran los siguientes resultados con una escala (hipotética) de dedicación al trabajo, con valores de 0 a 25:

R	G ₁	11	11	11.2	X ₁	16	18	21
R	G ₂	10.8	11	10.9	X ₂	15	14	11.8
R	G ₃	11.1	10.9	11.3	-	11	10.8	11.4

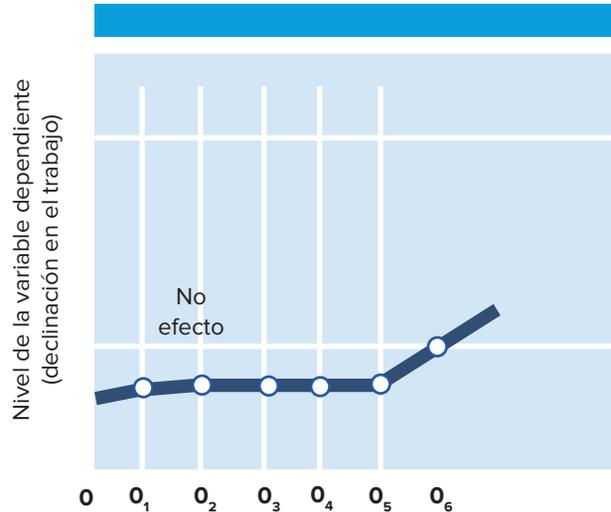
Podemos observar que X₁ tiene un efecto que aumenta con el paso del tiempo y X₂ produce un efecto en el corto plazo, pero tiende a desvanecerse. Esto podría graficarse del siguiente modo:



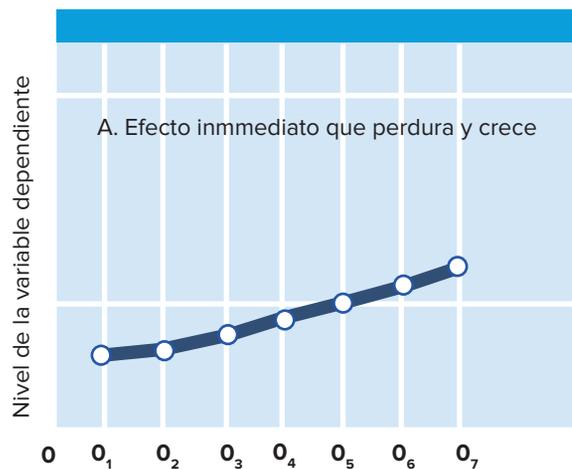


En los diseños experimentales de series cronológicas se realiza este tipo de gráficas, las cuales enriquecen la interpretación de la evolución de los grupos. Algunas de las diversas configuraciones que se presentan se muestran en las **figuras 5.3, 5.4 y 5.5**.

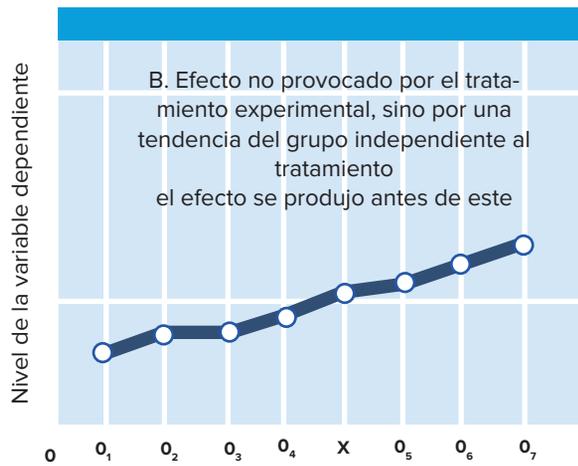
En este caso, debe observarse que no sea algún suceso ajeno el que provoque el efecto, en lugar de la manipulación de la variable independiente. Podría ocurrir que:



◆ **Figura 5.3** Ejemplo de un suceso ajeno a X que provoca el efecto.



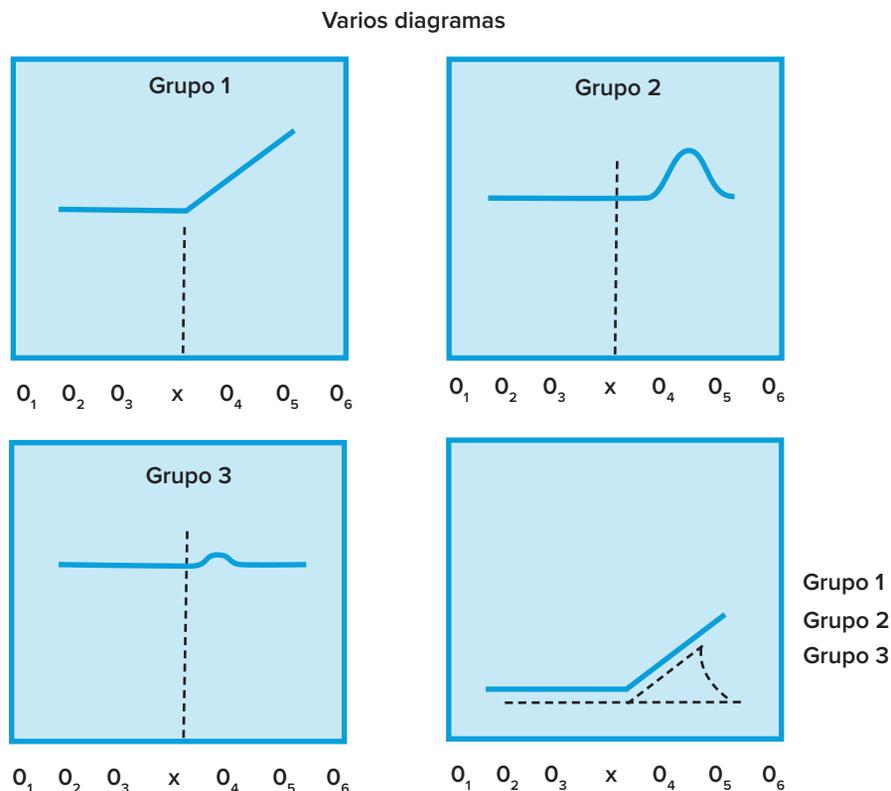
◆ **Figura 5.4** Ejemplo de un efecto que perdura.



◆ **Figura 5.5** Ejemplo de ausencia de efecto por el tratamiento.

Este diagrama se podría comparar con el del grupo de control o demás grupos para analizar lo que ocurre. Quizá se deba a que no se obtuvo una equivalencia real de los grupos al inicio del experimento.

Se podrían elaborar tantos diagramas o gráficas como grupos, o bien podrían agruparse los efectos provocados en los distintos grupos en una sola gráfica (varios diagramas en esta), como se observa en la **figura 5.6**.



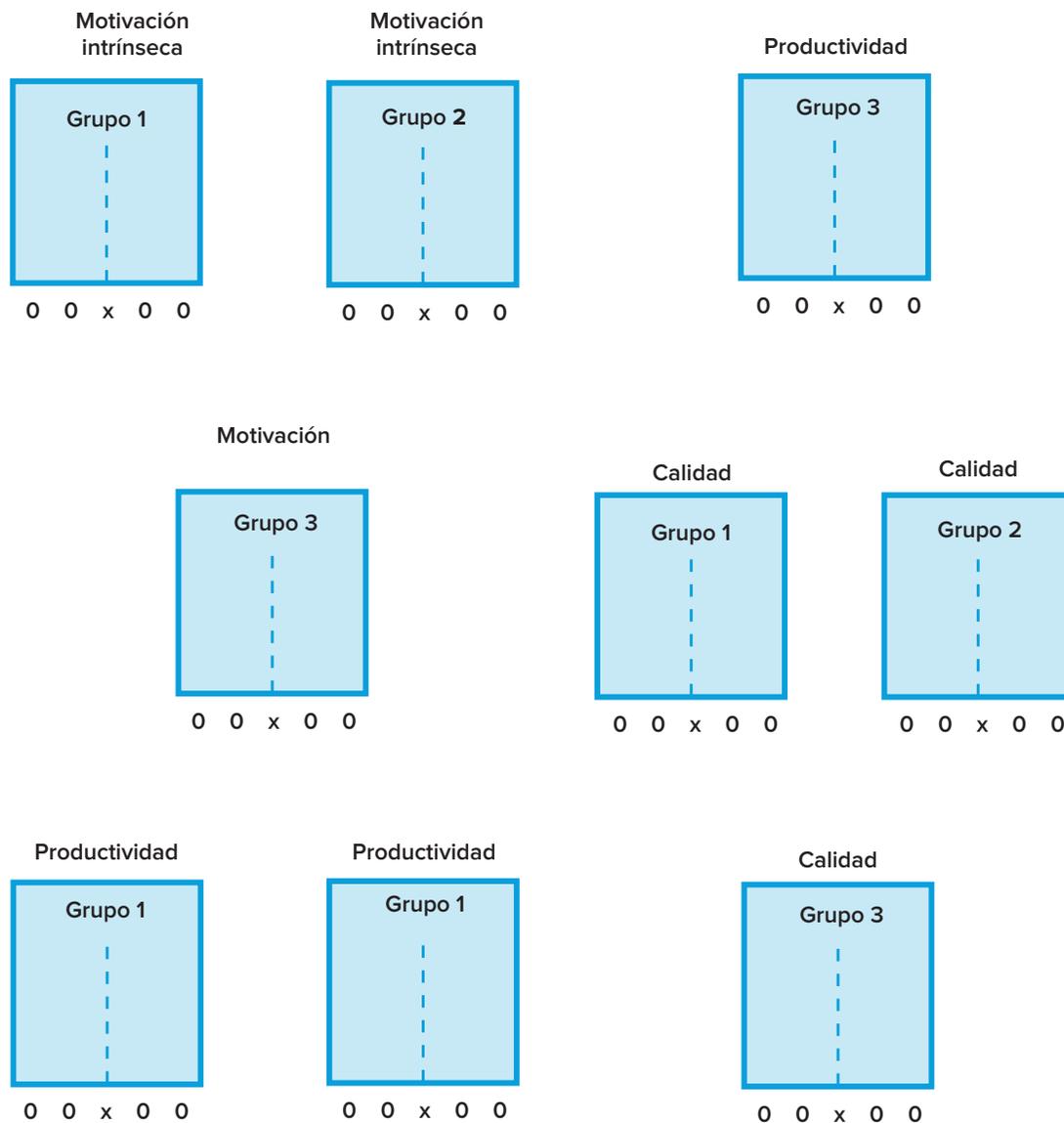
◆ **Figura 5.6** Ejemplo de gráficas de series cronológicas.



Ejemplo

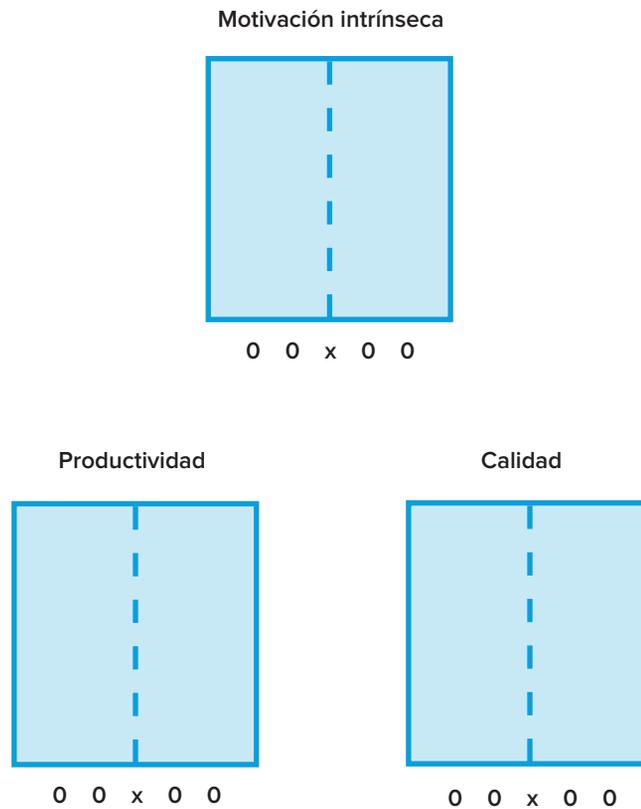
Si el tratamiento es un programa motivacional y las variables dependientes son la motivación intrínseca, la productividad y la calidad de la producción (con dos grupos experimentales y uno de control), se presentarían las opciones que se muestran en las figuras 5.7 y 5.8.

Diversos diagramas para varios grupos y variables independientes



◆ **Figura 5.7** Ejemplo de varios diagramas para cada variable dependiente.

Un solo diagrama por variable independiente



◆ **Figura 5.8** Ejemplo de un solo diagrama por variable dependiente.

Desde luego, si se mide más de una variable dependiente, en el primer caso se tendrá un diagrama para cada grupo por cada variable dependiente, mientras que en el segundo, un diagrama por cada variable dependiente.

En estos diseños de series cronológicas se controlan todas las fuentes de invalidación interna, siempre y cuando se lleve a cabo un seguimiento minucioso de los grupos, para asegurarse de que la única diferencia entre ellos sea la manipulación de la variable independiente. En algunos casos, la repetición de las mediciones influye sobre la variable dependiente (administración de pruebas múltiples), sobre todo en las pruebas donde el individuo participa activamente y sabe que está respondiendo a una medición (cuestionarios, entrevistas, tests estandarizados, aparatos o equipos, etc.), no tanto así en las medidas en las que el participante es más pasivo y no se encuentra consciente de qué se le mide (por ejemplo, la observación). De cualquier manera, en caso de que exista dicha influencia, se presentará de forma similar en todos los grupos (porque son equivalentes y el número de pruebas aplicadas es el mismo).

Para estos diseños se suelen utilizar diversas técnicas estadísticas complejas, lo cual depende del nivel de medición de las variables y del tipo de análisis e interpretación que se desee realizar, como análisis de regresión múltiple o análisis de cambio y series



de tiempo. Se recomiendan las siguientes fuentes para conocer tales análisis: sugerimos a Shmueli (2012), Cryer y Chan (2009); Hawkins (2008); Box, Jenkins y Reinsel (2008); Kollo y von Rosen (2005), y Kerlinger y Pedhazur (1997).

Diseños de series cronológicas con repetición del estímulo

Con cierta frecuencia, el investigador anticipa que el tratamiento o estímulo experimental no tiene efecto o este es mínimo si se aplica una sola vez, tal como sería hacer ejercicio físico un solo día (no se esperaría un cambio en la musculatura) o consumir vitaminas por una vez. También, en otras situaciones el investigador quiere conocer el efecto sobre las variables dependientes en cada ocasión que se aplica el estímulo experimental.

Por ejemplo, en técnicas de condicionamiento es común preguntarse cuántas veces se debe aplicar el reforzamiento a una conducta para condicionar la respuesta a un estímulo. En estos casos es posible repetir el tratamiento experimental y administrar una posprueba después de cada aplicación, para evaluar su efecto.

Los participantes se asignan al azar a los distintos grupos y a cada uno se le administra varias veces el tratamiento experimental que le corresponde. Algunos de estos diseños diagramados se muestran en la **figura 5.9**.

El mismo tratamiento se aplica dos veces al grupo experimental:

R	G ₁	O ₁	X ₁	O ₂	X ₁	O ₃
R	G ₂	O ₄	—	O ₅	—	O ₆

Cada tratamiento se aplica cuatro veces al grupo respectivo:

R	G ₁	O ₁	X ₁	O ₂	X ₁	O ₃	X ₁	O ₄	X ₁	O ₀₅
R	G ₂	O ₆	X ₂	O ₇	X ₂	O ₈	X ₂	O ₉	X ₂	O ₁₀
R	G ₃	O ₁₁	X ₃	O ₁₂	X ₃	O ₁₃	X ₃	O ₁₄	X ₃	O ₁₅
R	G ₄	O ₁₆	—	O ₁₇	—	O ₁₈	—	O ₁₉	—	O ₂₀

En algunos casos se podría prescindir de las pruebas, y el experimentador, por alguna justificación teórica o empírica, podría aplicar pospruebas a intervalos sistemáticos diferentes.

R	G ₁	X ₁	O ₁	X ₁	X ₁	X ₁	O ₂	X ₁	X ₁	X ₁	O ₃
R	G ₂	X ₂	O ₄	X ₂	X ₂	X ₂	O ₅	X ₂	X ₂	X ₂	O ₆
R	G ₃	—	O ₇	—	—	—	O ₈	—	—	—	O ₉

O bien, aplicar las pospruebas a intervalos irregulares (por alguna determinada razón):

R	G ₁	X ₁	X ₁	O ₁	X ₁	O ₂	X ₁	X ₁	X ₁	O ₀₃	X ₁	O ₀₄
R	G ₂	X ₂	X ₂	O ₅	X ₂	O ₆	X ₂	X ₂	X ₂	O ₀₇	X ₂	O ₀₈
R	G ₃	—	—	O ₉	—	O ₁₀	—	—	—	O ₁₁	—	O ₁₂

● **Figura 5.9** Ejemplo de diseños cronológicos con repetición del estímulo.

Un ejemplo de estos diseños sería el caso de un publicista que pretende analizar los efectos de un comercial televisivo en la preferencia del producto anunciado, en comparación con otras marcas, y postula la hipótesis de que una sola exposición al comercial no surtirá efecto alguno. Asimismo, son herramientas comunes para probar la eficacia de medicamentos o tratamientos psicológicos.

Las pruebas estadísticas usuales para validar estos diseños son las mismas que para las series cronológicas múltiples.

Diseños con tratamientos múltiples

A veces, el investigador desea analizar el efecto de la aplicación de los distintos tratamientos experimentales a todos los grupos o participantes. En estos casos es posible utilizar los diseños con tratamientos múltiples. La aplicación de tratamientos puede ser individual o grupal y asumir distintas configuraciones:

- a) **Varios grupos.** En este caso se trabaja con varios grupos a los cuales se asignan los sujetos o participantes al azar. A cada grupo se le aplican todos los tratamientos, cuya secuencia puede o no ser la misma para todos los grupos y es posible administrar una o más pospruebas (posteriores a cada tratamiento experimental). Dos posibles diagramas de estos diseños son los de la **figura 5.10**.

		Misma secuencia para los grupos						Secuencia diferente							
R	G1	X ₁	O ₁	X ₂	O ₂	X ₃	O ₃	R	G ₁	X ₁	O ₁	X ₂	O ₂	X ₃	O ₃
R	G2	X ₁	O ₄	X ₂	O ₅	X ₃	O ₆	R	G ₂	X ₂	O ₄	X ₃	O ₅	X ₁	O ₆
R	G3	X ₁	O ₇	X ₂	O ₈	X ₃	O ₉	R	G ₃	X ₃	O ₇	X ₂	O ₈	X ₁	O ₉

● **Figura 5.10** Diseños con tratamientos múltiples, varios grupos.

Secuencia diferente. El experimentador debe tener cuidado cuando interpreta las segundas pospruebas y mediciones subsecuentes, ya que puede existir una influencia diferente en los grupos provocada por las distintas secuencias de los tratamientos. En realidad, durante el experimento es muy probable que haya diferencias entre grupos, e incluso que al finalizar el experimento los resultados se deban, en buena medida, a la secuencia con que fueron administrados los tratamientos.

Los diseños experimentales con tratamientos múltiples y secuencia diferente en los grupos, así como los dos casos que vamos a ver a continuación, llegan a tener distintos efectos que deben analizarse con minuciosidad. Algunos tratamientos tienen efectos reversibles; en este caso no hay interferencia entre tratamientos, a la vez que las pospruebas son influidas solo por el tratamiento inmediato anterior (por ejemplo O₃, del diseño con secuencia diferente, se vería afectada por X₃, pero no por X₂ o X₁), lo cual facilita la interpretación. Sin embargo, a menudo los efectos no son reversibles, sino aditivos o interactivos; esto es, los resultados de una posprueba pueden ser influidos no solo por el tratamiento inmediatamente anterior, sino por los que antecedieron a este, y no es fácil saber cuánto se debió a X₁, cuánto a X₂ o X_k. Para ello, en el análisis debe incluirse la secuencia como factor.



- b) Un solo grupo. En situaciones en las que el experimento cuenta con un número reducido de participantes, es posible realizar un diseño con tratamientos múltiples y un solo grupo. No hay asignación al azar puesto que se tiene a un único grupo. La equivalencia se obtiene puesto que no hay nada más similar a un grupo que este mismo. El grupo hace las veces de “grupos experimentales” y de “control”. Este diseño se diagrama así:

$$G \text{ único } X_1 \ 0_1 \ X_2 \ 0_2 \ - \ 0_3 \ X_3 \ 0_4 \ - \ 0_5 \ X_k \ 0_k \dots$$

Cuando se considere conveniente, se utiliza como grupo de control, por ejemplo, antes de 03 y 05. Sin embargo, tal diseño está limitado a que los efectos de los tratamientos múltiples sean reversibles; de lo contrario no es un diseño experimental, sino cuasiexperimental. Si en estos diseños se introduce sistemáticamente y como variable independiente la secuencia de administración de los tratamientos, se convierten en factoriales (que se verán a continuación).

Con el fin de comprobar si las secuencias se relacionan con características de los grupos, algunos autores consideran que es útil introducir sistemáticamente otros elementos a los diseños, tales como presencia-ausencia de la preprueba (por ejemplo, el diseño de cuatro grupos de Solomon, que se vio en el capítulo siete del texto impreso), así como secuencias diferentes de tratamientos en varios grupos. Por ejemplo, se podría tener un diseño así:

Niños de 8o. año	G_1	X_1	0_1	X_2	0_2
	G_2	X_2	0_3	X_1	0_4
Niños de 6o. año	G_3	X_1	0_5	X_2	0_6
	G_4	X_2	0_7	X_1	0_8
Niños de 4o. año	G_5	X_1	0_9	X_2	0_{10}
	G_6	X_2	0_{11}	X_1	0_{12}

O un número de pospruebas diferentes en los grupos. Por ejemplo, con un diseño como el siguiente, se tienen diseños factoriales:

R	G_1	X_1	0_1	0_2
R	G_2	X_1	0_3	
R	G_3	X_2	0_4	0_5
R	G_4	X_2	0_6	
R	G_5	–	0_7	0_8
R	G_6	–	0_9	

Son factoriales, porque se manipulan como si fueran una variable independiente. Sin embargo, la experiencia ha demostrado a algunos profesores que los alumnos que ven

por primera vez el tema suelen desconcertarse si se analizan como diseños factoriales, a menos que estos se expliquen primero.

Las pruebas estadísticas que se utilizan en estos diseños son las mismas que se emplean para las series cronológicas y los diseños con repetición del estímulo.

Hasta aquí se han revisado diseños experimentales que manipulan una sola variable independiente (los tratamientos representan niveles de presencia o modalidades de manipulación de esta; X_1, X_2, X_3, X_k son variaciones de la misma variable independiente); sin embargo, como ya se señaló en la obra impresa, normalmente se miden diversas variables dependientes para optimizar el costo e incrementar el conocimiento, lo cual permite analizar los efectos de la variable independiente sobre cada dependiente. Por ejemplo, en un experimento con métodos educativos, en lugar de medir nada más el aprendizaje, es posible medir también la motivación del alumno, su integración al grupo, etc., lo cual se debería realizar en la preprueba y la posprueba. O bien, cuando se experimenta con un medicamento (digamos para controlar la hipertensión), se incluyen diversas variables dependientes, como el nivel de presión arterial, glucosa, peso, etc. Cuando se introduce como estímulo una nueva tecnología en una línea de producción, se consideran distintas variables dependientes tales como productividad, tiempo de procesamiento, parámetros de calidad, etcétera.

Asimismo, por lo regular se incluyen mediciones para verificar qué tanto funcionó la manipulación (verificaciones de la manipulación).

Diseños factoriales

Los diseños factoriales manipulan dos o más variables independientes e incluyen dos o más niveles o modalidades de presencia en cada una de ellas para analizar su efecto sobre una o más variables dependientes. Este enfoque se utiliza muy a menudo en la investigación del comportamiento, las ciencias de la salud y las diferentes ramas de la ingeniería. La construcción básica de un diseño factorial consiste en que todos los niveles de cada variable independiente sean tomados en combinación con todos los niveles de las otras variables independientes (Patience, 2013, *The SAGE glossary of the social and behavioral sciences*, 2009t y Cramer, 2003).

Diseño factorial 2×2

El diseño factorial más simple manipula (hace variar) dos variables, cada una con dos niveles. A este diseño se le conoce como “diseño factorial 2×2 ”, en donde el número de dígitos indica el número de variables independientes:

2	×	2
Un dígito (primera variable independiente)		Otro dígito (segunda variable independiente)

Número de grupos resultantes o del experimento

El valor numérico de cada dígito indica el número de niveles o modalidades de la variable independiente en cuestión. En este caso es 2, lo cual quiere decir que cada una de



las variables tiene dos niveles. Como mencionan Wiersma y Jurs (2008), no es necesario que los valores numéricos sean los mismos para todas las variables independientes. En teoría, puede haber cualquier número de variables independientes con cualquier número de niveles cada una. Por ejemplo, el diseño factorial $2 \times 2 \times 3$ indica que hay tres variables independientes, la primera y la segunda con dos niveles, mientras que la tercera tiene tres niveles. El diseño factorial $4 \times 5 \times 2 \times 3$ indica una variable independiente con cuatro niveles, otra con cinco, otra más con dos y una última con tres.

Un ejemplo de un diseño factorial 2×2 sería que las variables independientes fueran “método de enseñanza” y “género”, y la variable dependiente el “aprovechamiento escolar”. La primera tiene dos variantes: “método de enseñanza presencial” y “método de enseñanza en línea”. La segunda, con las modalidades “masculina” y “femenina”. Otro ejemplo estaría constituido por las variables independientes “administración del medicamento” (con los niveles: suministro y placebo) y “condición de fumador” (fumador-no fumador), y la variable dependiente “control de la hipertensión arterial”.

Un diseño factorial $2 \times 3 \times 4$ (tres variables independientes) podría incluir: “tipo de medicamento” (medicamento A-medicamento B), “dosis diaria” (30 mg, 50 mg y 90 mg) y “edad” (menos de 30 años, 31-40 años, 41-50 años, más de 50 años). La variable dependiente podría ser “disminución de síntomas de la enfermedad” (Wright y London 2009).

También podríamos considerar un diseño 3×4 que experimentara la “resistencia al fuego” (variable dependiente), así como tres tipos de materiales (variable independiente 1) y cuatro clases de grosor (variable independiente 2).

Otros diseños factoriales

El número de grupos o condiciones que se forman en un diseño factorial es igual a todas las posibles combinaciones que surgen cuando se cruzan los niveles de una variable independiente con los niveles de las otras variables del mismo tipo. Así, en un diseño 2×2 tendremos cuatro grupos ($2 \text{ por } 2 = 4$); en un diseño 3×2 resultarán seis grupos, y en un diseño $3 \times 3 \times 3$, 27 grupos. Debe observarse que el resultado de la multiplicación es el número de grupos resultante. En estos diseños, el número de grupos aumenta con rapidez a la par con el incremento del número de variables independientes o niveles (exponencialmente). Veámoslo:

2×2	$= 4$	}	Número de grupos resultantes o del experimento
2×3	$= 6$		
4×3	$= 9$		
3×4	$= 12$		
$3 \times 2 \times 2$	$= 12$		
$3 \times 3 \times 4$	$= 26$		

Ello se debe a que los niveles tienen que tomarse en cuenta respecto de todas sus posibles combinaciones.

Cheng (2014) y Wiersma y Jurs (2008) señalan que en los diseños factoriales, al menos una de las variables independientes debe ser propiamente experimental; las

demás pueden ser variables orgánicas o cuya variación no es provocada para fines del experimento, introducidas en el diseño con fines de control [por ejemplo, género, edad, año, escolaridad, inteligencia en intervalos, etcétera].

		Diseño 2 × 2	
Variable independiente B		Variable independiente A	
		A ₁	A ₂
	B ₁	A ₁ B ₁	A ₂ B ₁
	B ₂	A ₁ B ₂	A ₂ B ₂

● **Figura 5.11** Un diseño factorial 2 × 2.

Con objeto de simplificar la forma en que se diagraman los diseños factoriales, acudiremos a la simbología que comúnmente se utiliza.² Para designar a las variables independientes se usan letras (A, B, C, ... K) y para los niveles, números (1, 2, 3, ... K); las combinaciones de letras y números que aparecen en las casillas (o celdas) representan las mezclas de niveles de las variables independientes. Cada celda es un grupo o condición. En la **figura 5.11** se diagrama un diseño factorial 2 × 2. Otro ejemplo sería un diseño factorial 2 × 4 × 3 (**figura 5.12**).

		A						
			A ₁			A ₂		
	C	C ₁	C ₂	C ₃	C ₁	C ₂	C ₃	
B	B ₁	A ₁ B ₁ C ₁	A ₁ B ₁ C ₂	A ₁ B ₁ C ₃	A ₂ B ₁ C ₁	A ₂ B ₁ C ₂	A ₂ B ₁ C ₃	
	B ₂	A ₁ B ₂ C ₁	A ₁ B ₂ C ₂	A ₁ B ₂ C ₃	A ₂ B ₂ C ₁	A ₂ B ₂ C ₂	A ₂ B ₂ C ₃	
	B ₃	A ₁ B ₃ C ₁	A ₁ B ₃ C ₂	A ₁ B ₃ C ₃	A ₂ B ₃ C ₁	A ₂ B ₃ C ₂	A ₂ B ₃ C ₃	
	B ₄	A ₁ B ₄ C ₁	A ₁ B ₄ C ₂	A ₁ B ₄ C ₃	A ₂ B ₄ C ₁	A ₂ B ₄ C ₂	A ₂ B ₄ C ₃	

● **Figura 5.12** Diseño 2 × 4 × 3.

En la figura 5.12 se puede observar que todas las posibles combinaciones de los niveles entre A, B y C están presentes; además, ninguna es exactamente igual a la otra. Cada combinación representa una celda o un grupo. Si las tres variables se manipulan

² Por ejemplo Wiersma y Jurs (2008), Ramsey (2006) y Christensen (2006).



deliberadamente (recordemos que al menos una debe serlo para que hablemos de experimento), los participantes tienen que asignarse al azar a todas las celdas o grupos. Si dos variables se manipulan intencionalmente (por ejemplo, B y C), los sujetos de cada nivel de la variable restante (A) serán asignados aleatoriamente a las casillas que les correspondan. Veámoslo con un ejemplo. Si A = género (A_1 , masculino; A_2 , femenino), B = violencia televisada (B_4 , elevada; B_3 , mediana; B_2 , baja; y B_1 , nula) y C = orientación paternal sobre el programa visto (C_1 , por parte de ambos padres; C_2 , del padre; y C_3 , de la madre). Si hay 120 niños y 120 niñas, los niños (A_1) se asignarían al azar a las celdas donde A_1 está presente (10 niños en cada celda), y las niñas (A_2) a las 12 casillas restantes (donde A_2 está presente, 10 niñas por celda). Si una sola variable es la que se manipula deliberadamente (por ejemplo, C), los participantes de los niveles combinados de las otras dos variables se asignan al azar a los niveles de aquella (C_1 , C_2 y C_3 en el ejemplo). Los sujetos $A_1 B_1$ serían asignados aleatoriamente a C_1 , C_2 y C_3 , igual los sujetos $A_1 B_2$, $A_1 B_3$, etcétera.

Cuando los diseños son factoriales es posible agregar un grupo de control o varios (que no se expongan a la variable o las variables manipuladas deliberadamente, **figura 5.13**).

Diseño 2×2

	A_1	A_2	
B_1	$A_1 B_1$	$A_2 B_1$	
B_2	$A_1 B_2$	$A_2 B_2$	
			Control

● **Figura 5.13** Diseño factorial con grupo de control.

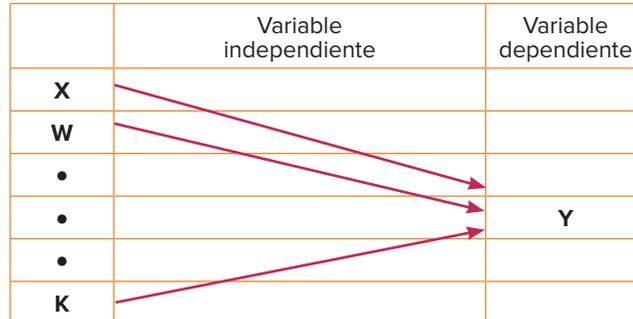
Utilidad de los diseños factoriales

Los diseños factoriales son sumamente útiles porque permiten al investigador evaluar los efectos de cada variable independiente sobre la dependiente por separado, así como los efectos de las variables independientes de manera conjunta. Por medio de estos diseños se observan los efectos de interacción entre las variables independientes.

En términos de Wiersma y Jurs (2008), la interacción es un efecto producido sobre la variable dependiente, de tal manera que el efecto de una variable independiente no permanece constante en los niveles de la otra u otras independientes. El efecto de interacción está presente si el efecto conjunto de las variables independientes no es igual a sus efectos por separado (aditivos). Ello significa que el efecto de una variable independiente por sí mismo no es igual que cuando se toma en combinación con los niveles de otra variable independiente. Por ejemplo, si el alto contenido de violencia televisada afecta solo cuando hay orientación sobre el programa por parte de la madre, pero no cuando dicha orientación está a cargo del padre o de ambos.

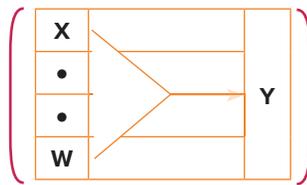
Así, hay dos tipos de efectos que es posible evaluar en los diseños factoriales: los efectos de cada variable independiente (llamados efectos principales) y los efectos de interacción entre dos o más variables independientes (si se consideran cuatro variables,

por ejemplo, pueden interactuar dos entre sí y otras dos entre sí, o pueden interactuar tres o las cuatro variables independientes). Los diseños factoriales responden a estructuras entre variables que se esquematizan en la siguiente forma:

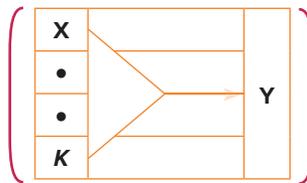


• Figura 5.14 Ejemplo de estructura de un diseño factorial.

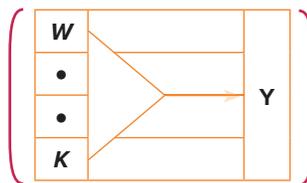
Asimismo, se analizan el efecto de X sobre Y ($X \rightarrow Y$), de W sobre Y ($W \rightarrow Y$) y el de ...K sobre Y ($K \rightarrow Y$); el efecto conjunto de X y W sobre Y es



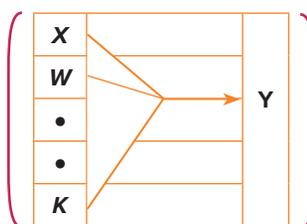
y el efecto conjunto de X y ...K sobre Y



Por su parte, el de W y ...K sobre Y sería



Finalmente, el efecto conjunto de todas las variables independientes sobre Y (tres variables) es el siguiente:





Ejemplo

Supongamos una investigación cuya hipótesis es: "A mayor exposición por parte de jóvenes universitarios a videos musicales con alto contenido sexual, habrá una mayor predisposición para establecer contacto heterosexual".

Posteriormente se diseña un experimento para someterla a prueba. La variable independiente es la exposición a contenidos sexuales (por medio de la televisión) y la dependiente es la predisposición para establecer contacto sexual (relaciones sexuales). Se decide agregar otra variable independiente: género. En consecuencia, se ha elaborado un diseño factorial con dos variables independientes. La exposición tendría tres niveles: a) contenido sexual elevado y manifiesto, b) contenido sexual moderado y c) contenido "romántico".

En consecuencia, cada video que se elabore tendrá tres versiones: la música es la misma, los modelos, la duración, la historia y el contexto también. La única diferencia es el tratamiento sexual de los contenidos verbales y no verbales. El género implicaría dos niveles: masculino y femenino.

El esquema del diseño se indica en la **figura 5.15**.

		Exposición al contenido sexual		
		Elevada	Moderada	Mínima
Género	Masculino			
	Femenino			

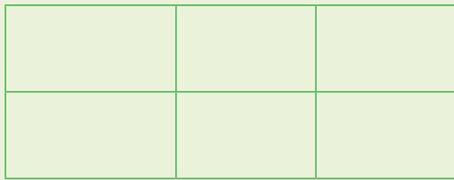
● **Figura 5.15** Diseño factorial 2 x 3 (correspondiente al ejemplo).

Christensen (2006) desarrolla una excelente explicación de los efectos principales e interactivos en los diseños factoriales, la cual sirve de base para la exposición de dichos efectos que se presenta a continuación.

Supongamos que, una vez que los grupos se han expuesto a los videos, se les aplica una posprueba que mide su predisposición para establecer contacto sexual y se obtiene el promedio de cada grupo (asimismo, pensemos que los resultados que lleguen a obtenerse en esta prueba oscilaran entre 0 y 30, donde un valor mayor indica una más alta predisposición). Analizamos varias configuraciones posibles de resultados (**figura 5.16**).

Interpretación y posibles configuraciones

En la configuración A, no hay diferencias por género (cada casilla en cada nivel de género es igual a la del otro nivel). En cambio, sí las hay entre los niveles de exposición en ambos géneros. Los efectos principales (es decir, los de cada variable por separado) se observan comparando los promedios de los marginales de los niveles de cada variable (vea la interpretación de efectos de la **figura 5.16**, primera parte).



Marginal

Marginal

Si hay diferencias significativas entre los promedios de los marginales, hubo efecto de la variable "horizontal" (en el ejemplo, el género).

Marginal Marginal Marginal

Si hay diferencias significativas entre los promedios de los marginales, hubo efecto de la variable "vertical" (en el ejemplo, exposición al contenido sexual).

Interpretación de efectos

Promedios de marginales	25	15	7	15.7	No efecto
	25	15	7	15.7	
	25	15	7		

Efecto

A

	Elevada	Moderada	Mínima
Masculino	25	15	7
Femenino	25	15	7

La variable exposición tiene efecto; la variable género no tiene efecto. No hay efecto de interacción.

B

	Elevada	Moderada	Mínima
Masculino	18	18	18
Femenino	8	8	8

La variable exposición no tiene efecto; la variable género tiene efecto. No hay efecto de interacción.

C

	Elevada	Moderada	Mínima
Masculino	26	18	10
Femenino	10	18	25.9

Ninguna de las variables tiene efecto por sí misma, pero hay un efecto de interacción.

D

	Elevada	Moderada	Mínima
Masculino	23	18	13
Femenino	18	13	8

Las variables tienen efecto, pero no hay efecto de interacción.

● **Figura 5.16** Diagramas del ejemplo de diseño factorial.

E				F			
	Elevada	Moderada	Mínima		Elevada	Moderada	Mínima
Masculino	29	23	12	Masculino	8	7.9	8.1
Femenino	20	12	12	Femenino	7.9	8	8

Las dos variables tienen efecto y hay efecto de interacción.

No hay ningún tipo de efecto.

● **Figura 5.16** Diagramas del ejemplo de diseño factorial (*continuación*).

En la configuración B, no hay diferencias por exposición, pero sí por género (el marginal masculino es mayor).

En la configuración C, las diferencias entre las celdas se explican porque las dos variables interactúan (cuando el género es masculino y la exposición elevada, y cuando el género es femenino y la exposición mínima, se obtiene un valor alto; cuando el género es masculino y la exposición mínima, y cuando el género es femenino y la exposición elevada, se obtiene otro valor bajo; finalmente, cuando ambos géneros se exponen moderadamente se obtiene un valor distinto de las demás celdas). El máximo efecto se encuentra en “masculino” y “elevada” y “femenino” y “mínima” (26 y 25.9). No se registran efectos principales.

En la configuración D, hay cambios verticales y horizontales provocados por efectos principales, pero no efecto de interacción (ambas variables tienen efecto por sí mismas, únicamente).

En la configuración E, hay efectos principales (cada variable por sí misma afecta significativamente) y también efecto de interacción (este se alcanza si la diferencia entre las medias de los niveles de variación de una variable independiente cambian en función de los niveles de variación de la otra variable independiente, como también ocurrió en la configuración C).

En la configuración F, las diferencias entre todas las celdas es prácticamente nula: no hay ninguna clase de efecto.

Métodos estadísticos de los diseños factoriales

Los métodos estadísticos más usuales que se aplican a estos diseños son el análisis de varianza factorial (ANOVA) y el análisis de covarianza (ANCOVA), con la variable dependiente medida en intervalos, y la Chi cuadrada para múltiples grupos, con esa variable medida nominalmente.

Por último, a estos diseños se les pueden agregar más variables dependientes (tener dos o más) y se convierten en diseños multivariados experimentales que utilizan como método estadístico el análisis multivariado de varianza (MANOVA).

¿Qué otros experimentos existen? Cuasiexperimentos

Ya se comentó en el capítulo 7 del libro que los diseños cuasiexperimentales también manipulan deliberadamente una o más variables independientes para observar su efecto y relación con una o varias dependientes, solo que trabajan con “grupos intactos”, formados por motivos ajenos al experimento: cuando se opta por los diseños cuasiexperimentales los participantes no se asignan al azar a los grupos ni se emparejan, sino que dichos grupos se integran antes del experimento (Thyer, 2012 y Stone-Romero, 2006). Por ejemplo, los grupos del experimento son cuatro de diferentes áreas de una empresa que estaban conformados con anterioridad al experimento (las razones por las que cada participante se integró a cada área no se relacionan con la realización del experimento). Veámoslo gráficamente:

Cuasiexperimento Experimento en el que los sujetos no se asignan al azar a los grupos ni se emparejan, porque tales grupos ya existían.



Grupo A	(50 empleados del departamento de producción)	∅	Grupo experimental con X1
Grupo B	(26 empleados de la gerencia de recursos humanos)	∅	Grupo experimental con X2
Grupo C	(37 empleados de la gerencia de administración y finanzas)	∅	Grupo experimental con X3
Grupo D	(29 empleados del departamento de aseguramiento de la calidad)	∅	Grupo de control

De igual manera, pueden ser pacientes de dos hospitales, uno el grupo experimental y el otro el de control; o bien, tres líneas de producción integradas por distintos colaboradores.

Problema de los diseños cuasiexperimentales

Estos diseños se utilizan cuando no es posible asignar en forma aleatoria los participantes a los grupos que recibirán los tratamientos experimentales. La falta de aleatorización introduce posibles problemas de validez interna y externa. Como comenta Weiss (1990, p. 89):

[...] estos diseños deben luchar con la selección como fuente posible de interpretación equivocada, lo mismo que con la interacción de la selección y otros factores; así como, posiblemente, con los efectos de la regresión.

Asimismo, diversos elementos pudieron operar en la formación de los grupos (que no están bajo el control del investigador), los cuales impiden afirmar que sean representativos de poblaciones más amplias (Holmes, 2013). De este modo, dado que su validez es menor que la de los experimentos “puros”, reciben el nombre de **cuasiexperimentos**.

A causa de los problemas potenciales de validez interna, en estos diseños el investigador debe intentar establecer la semejanza entre los grupos, lo cual requiere considerar las características o variables que estén relacionadas con las variables estudiadas (Babbie, 2014



y Wiersma y Jurs, 2008). Por ejemplo, si grupos intactos de trabajadores se involucran en un experimento sobre motivación o la introducción de ciertas normas de calidad, el turno probablemente tenga que ser introducido como una constante (grupos intactos, que sean todos del mismo turno) o como otra variable independiente (de control). Además, el investigador debe recabar evidencia de que los grupos son equiparables en salario, productividad, competencia, antigüedad en la organización y, en general, en todo lo que genere diferencias significativas entre ellos. Cuanta mayor información se obtenga sobre los grupos, mayores bases se tendrán para establecer su semejanza. En algunos casos se debe observar si la proporción de mujeres y hombres en los grupos es semejante, si la edad promedio es similar, si los grupos no fueron constituidos con base en un criterio que pudiera afectar (por ejemplo, formación de los salones por inteligencia) y si a los grupos en el pasado no les ha ocurrido algo que llegara a influir en los resultados. Además, como mencionan Campbell y Stanley (1966, p. 70):

...precisamente porque hay falta de control experimental total, es imprescindible que el investigador conozca a fondo cuáles son las variables particulares que su diseño específico no controla. Así, estará más pendiente de su posible influencia y tendrá mejores elementos para evaluarla.

La ausencia de asignación al azar obliga al investigador a poner especial atención cuando interpreta los resultados para no caer en conclusiones erróneas. Las limitaciones deben identificarse con claridad, la equivalencia de los grupos debe discutirse y la posibilidad de generalizar los resultados, así como la representatividad, deben argumentarse sobre una base lógica (Thyer, 2012, Wiersma y Jurs, 2008 y Reichardt, 2004b).

Los **cuasiexperimentos** difieren de los experimentos “puros” en la equivalencia inicial de los grupos (los primeros trabajan con grupos intactos y los segundos utilizan un método para hacer equivalentes a los grupos). Sin embargo, esto no quiere decir que sea imposible realizar un cuasiexperimento en el cual las variables relevantes para el estudio sean equiparables. Si así fuera, los cuasiexperimentos ya se hubieran desechado como diseños de investigación. Más bien quiere decir que, en algunos casos, los grupos pueden diferir significativamente o no ser equiparables; de este modo, el investigador debe analizar si los grupos son o no comparables desde el inicio. En caso negativo, el investigador no debe realizar la investigación con fines explicativos, esto es, limitarla a propósitos descriptivos y/o correlacionales.

Una recomendación útil para consolidar la validez interna de esta clase de diseños experimentales es asignar al azar los tratamientos o estímulos.

Tipos de diseños cuasiexperimentales

Con excepción de la diferencia que acabamos de mencionar, los cuasiexperimentos son muy parecidos a los experimentos “puros”. Por lo tanto, podemos decir que hay tantos diseños cuasiexperimentales como experimentales “puros”, solo que no se recurre a la asignación al azar ni al emparejamiento. Por lo demás son iguales, pues la interpretación es similar, las comparaciones y los análisis estadísticos son los mismos (salvo que a veces se consideran las pruebas para datos no correlacionados y relaciones no lineales). Es por ello que nos limitaremos a ver solo algunos de los diseños cuasiexperimentales (el resto puede

ser deducido de sus correspondientes diseños experimentales “puros”, con solo quitarles la “R” de asignación al azar) y se comentarán brevemente porque se insiste en que las comparaciones, interpretaciones y los análisis son prácticamente equiparables.³ Consideramos que no sería adecuado volver a explicar dichas comparaciones, interpretaciones y análisis.

1. *Diseño con posprueba únicamente y grupos intactos.* Este primer diseño utiliza dos grupos, pero solo uno recibe el tratamiento experimental. Los grupos son comparados en la posprueba para analizar si el tratamiento experimental tuvo un efecto sobre la variable dependiente (o_1 con o_2). El diseño puede diagramarse del siguiente modo, en el cual, tal como se ve, no hay asignación al azar ni emparejamiento:

$$\begin{array}{rcc} G_1 & \times & o_1 \\ G_2 & - & o_2 \end{array}$$

Observe que si los grupos *no* son equiparables entre sí, las diferencias entre las pospruebas de ambos grupos se pueden atribuir a la variable independiente o a otras razones, pero lo peor es que el investigador quizá no se dé cuenta de ello. Por ejemplo, supongamos que se lleva a cabo un cuasiexperimento para analizar el efecto de la realimentación que los médicos dan a sus pacientes (respecto de su conducta en el tratamiento prescrito) sobre la obediencia o apego al tratamiento. En este caso se podría partir de la siguiente hipótesis: “Los pacientes que reciban mayor realimentación de parte de sus médicos, acerca de cómo se comportan en el tratamiento prescrito, se apegarán más a este”. Es decir, los médicos que informen más a sus pacientes sobre su observancia del tratamiento prescrito propiciarán en estos un mayor deseo de continuar con la terapia. La cuestión es motivar al paciente. Entonces, el investigador toma dos grupos de pacientes. Uno de ellos recibe realimentación sobre su conducta en el tratamiento prescrito y el otro, no. Posteriormente se evalúa en qué medida se apega cada grupo, en lo sucesivo, al tratamiento. Supongamos que obtenemos el siguiente resultado: $o_1 > o_2$ (el grupo experimental se apega más); entonces, deducimos que la hipótesis fue confirmada. Pero, para estar seguros de ello, debemos analizar con mucho cuidado que sea posible comparar a los grupos.

Imaginemos que el grupo experimental estaba formado por pacientes que asisten a un hospital donde con frecuencia se imparten pláticas motivadoras para que los enfermos sigan los tratamientos prescritos, mientras que el grupo de control estaba integrado por personas que asisten a un hospital donde no se le asigna importancia a ello. ¿A qué razón se le podrían atribuir con certeza los resultados? ¿A la manipulación de la variable independiente, a que los grupos de pacientes provienen de diferentes hospitales, a ambos factores, a la sensibilización previa o a algún otro? Como los grupos no son razonablemente equiparables, no tendríamos la certeza de cuál fue la causa o qué tanto contribuyeron los diversos factores involucrados. En casos como este podemos decir que existe un problema de validez interna.

³ Si al lector le surge alguna duda respecto de qué comparaciones, interpretaciones y análisis pueden llevarse a cabo en un diseño cuasiexperimental, le recomendamos que revise el diseño experimental “puro” correspondiente. Vale recordar que la diferencia es que en el cuasiexperimental no hay aleatorización y los grupos pueden estar no correlacionados.



También podría ser que el grupo experimental estuviera compuesto por pacientes que, desde antes del experimento, tuvieran una motivación elevada para apearse a tratamientos médicos; o tal vez hayan operado otras causas que provocaran diferencias iniciales entre los grupos. Por ello, es importante que los grupos sean inicialmente comparables y que durante el experimento no ocurra algo que los haga diferentes, con excepción de la presencia-ausencia del tratamiento experimental (por ejemplo, misma enfermedad y tratamiento, hospital, médico que los atiende, instrucciones y lugar, equivalencia como grupos en género, edad, avance de la enfermedad, etc.; nada más imaginemos que el grupo experimental, en promedio, está “más enfermo” que el de control, y los pacientes lo saben; también llega a suceder que los más enfermos se apeguen más al tratamiento). El criterio de los experimentos “puros” de mantener la igualdad de los grupos (salvo la manipulación de la variable independiente) se aplica por igual a los cuasiexperimentos.

Puede extenderse el diseño para incluir más de dos grupos. Se tienen así diferentes tratamientos experimentales o niveles de manipulación. Su formato general sería:

G_1	X_1	0_1
G_2	X_2	0_2
G_3	X_3	0_3
•	•	•
•	•	•
G_k	X_k	0_k
G_{k+1}	—	0_{k+1}

El último grupo es de control. Un ejemplo de este diseño serían cuatro grupos escolares de un mismo semestre y carrera (licenciatura) en una universidad, como grupos del cuasiexperimento. Veámoslo esquemáticamente en la **figura 5.17**.

Universidad del Centro Escuela de Psicología Tercer semestre		
Grupo	X_1	0_1
Grupo	X_2	0_2
Grupo	X_3	0_3
Grupo	—	0_4

● **Figura 5.17** Diagrama de un ejemplo del diseño cuasiexperimental con posprueba únicamente.

Recuerde que los grupos son intactos (no se generan) y ya se habían constituido por motivos diferentes al cuasiexperimento (en este caso, la elección de estudiar una carrera y la asignación de alumnos a los grupos por parte de la escuela de psicología).

Los tratamientos experimentales podrían ser métodos educativos.

Otro ejemplo sería tener tres plantas de manufactura y dos sistemas de calidad o normas ISO experimentales (la tercera planta constituiría el grupo testigo).

Peña, Hancock y Merola (2009) efectuaron un cuasiexperimento que demostró que los individuos que utilizan avatares de color negro desarrollan intenciones más agresivas y menor cohesión grupal que quienes usan avatares de color blanco.⁴ El estudio apoya el efecto de proteus, el cual indica que en un ambiente virtual el tipo de vestimenta y caracterización del avatar puede tener efectos en la activación de pensamientos negativos (como la agresión) y la inhibición de pensamientos prosociales (por ejemplo, cohesión y afiliación). Incluso estos efectos podrían tener implicaciones para quienes en la vida real desempeñan papeles mediatizados por un ambiente virtual, como los pilotos de aviones de guerra.

2. Diseño con prueba-posprueba y grupos intactos (uno de ellos de control). Este diseño es similar al que incluye posprueba únicamente y grupos intactos, solo que en este caso a los grupos se les administra una preprueba, la cual puede servir para verificar la equivalencia inicial entre ellos (si son equiparables no debe haber diferencias significativas entre las prepruebas de los grupos). Su esquema más sencillo sería el siguiente:

$$\begin{array}{cccc} G_1 & 0_1 & X & 0_2 \\ G_2 & 0_3 & - & 0_4 \end{array}$$

Este enfoque puede extenderse a más de dos grupos (niveles o modalidades de manipulación de la variable independiente), lo cual se esquematizaría así:

G_1	0_1	X_1	0_2
G_2	0_3	X_2	0_4
G_3	0_5	X_3	0_6
•	•	•	•
•	•	•	•
G_k	0_{2k-1}	X_k	0_{2k}
G_{k+1}	0_{2k-1}	-	$0_{2(k+1)}$

Las posibles comparaciones entre las mediciones de la variable dependiente y las interpretaciones son las mismas que en el diseño experimental de preprueba-posprueba con grupo de control, solo que, en este segundo diseño cuasiexperimental, los grupos son intactos y en la interpretación de resultados debemos tomarlo en cuenta. Recuerde todo lo que se ha dicho de la probable no equivalencia de los grupos. Este aspecto se aplica a todos los diseños cuasiexperimentales.

3. Diseños cuasiexperimentales de series cronológicas

En ocasiones el investigador pretende analizar efectos a mediano y largo plazos o los efectos de la administración del tratamiento experimental varias veces, pero no cuen-

⁴ Recordemos una vez más que en este contexto un avatar es un personaje o representación de un usuario para participar en un juego o tarea.



ta con la posibilidad de asignar de manera aleatoria los casos, sujetos o participantes a los grupos del experimento. En estas condiciones, pueden utilizarse los diseños cuasiexperimentales, salvo que los grupos son intactos. En ambas situaciones se aplican mediciones repetidas de la variable dependiente y se inserta el tratamiento experimental entre dos de esas mediciones en, al menos, un grupo, mientras que al otro grupo no se le aplica ningún tratamiento en el periodo de “experimentación”. Sin embargo, desde la perspectiva de la literatura clásica sobre experimentos (vea Campbell y Stanley, 1966) se reconoce como cuasiexperimento a un diseño que no tiene grupo de control, tema que trataremos a continuación.

Domina (2009) implementó un diseño de esta naturaleza para evaluar la eficacia de los programas de asesoría, vinculación y orientación a alumnos que transitan de la educación media a la educación superior. El grupo experimental estuvo constituido por alumnos de escuelas que participaron en algún programa de este tipo y el grupo de control por estudiantes de instituciones que no contaban con un esfuerzo similar. El efecto se midió durante su desempeño en la universidad. Obviamente, no había posibilidad de asignación aleatoria. Al final, se eligieron casos similares en cada condición, una especie de “emparejamiento”.

Series cronológicas de un solo grupo

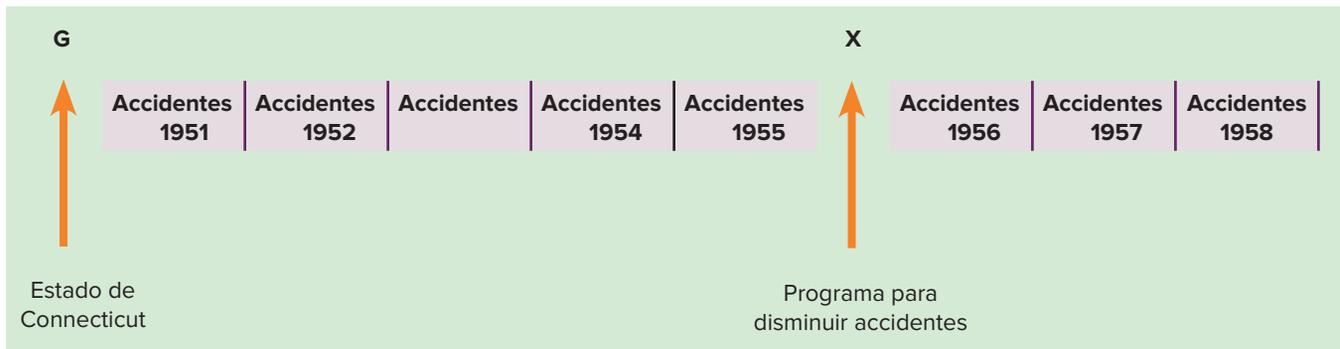
A un único grupo se le administran varias prepruebas, después se le aplica el tratamiento experimental y finalmente varias pospruebas. El diseño se diagrama así:

$$G \quad 0_1 \quad 0_2 \quad 0_3 \quad X \quad 0_4 \quad 0_5 \quad 0_6$$

El número de mediciones está sujeto a las necesidades específicas de la investigación que realizamos. Idealmente, puede convertirse en una serie cronológica con varios grupos (tema que se revisa más adelante).

Un ejemplo muy difundido de este diseño lo constituyó la evaluación de un programa que tenía por objeto disminuir la velocidad en carreteras del estado de Connecticut, Estados Unidos (Campbell, 1975). Los investigadores recolectaron informes y datos de accidentes de tránsito correspondientes a varios años anteriores y posteriores a la implantación del programa. Descubrieron que después del programa el número de accidentes disminuyó. Sin embargo, como las distintas mediciones habían mostrado una pauta ascendente y descendente inestable durante varios años, no se podía tener la certeza de que el programa hubiese sido la razón del descenso del número de accidentes (Weiss, 1990). En consecuencia, fue necesario comparar las estadísticas de Connecticut con las de otros cuatro estados vecinos en los que no se habían efectuado los cambios en los reglamentos de tránsito propuestos por el programa del mencionado estado. Dichas entidades actuaron como grupos de control. Finalmente, se observó que en los otros estados no se había registrado una disminución equivalente del número de accidentes. Las comparaciones dieron pie para concluir que el programa había generados los efectos deseados (Campbell, 1975 y Glass, 1968).

En su primera etapa, esta investigación cuasiexperimental utilizó las series cronológicas de un solo grupo (**figura 5.18**).



● **Figura 5.18** Ejemplo de un cuasiexperimento en su primera etapa.

Otro ejemplo de este diseño sería medir las ventas de un producto durante varios meses, introducir una campaña publicitaria para ese producto y, después, medir durante meses el nivel de ventas. Además, de nuevo, en una situación ideal, los grupos testigos podrían ser otros productos de la competencia sin campañas similares.

Estos diseños suelen utilizarse cuando se introducen innovaciones tecnológicas en empresas, probar resistencia y durabilidad de procesos de construcción, etcétera. También podrían emplearse para evaluar el efecto de reformas hacendarias, políticas públicas, medidas epidemiológicas (como una campaña de vacunación) o programas sociales.

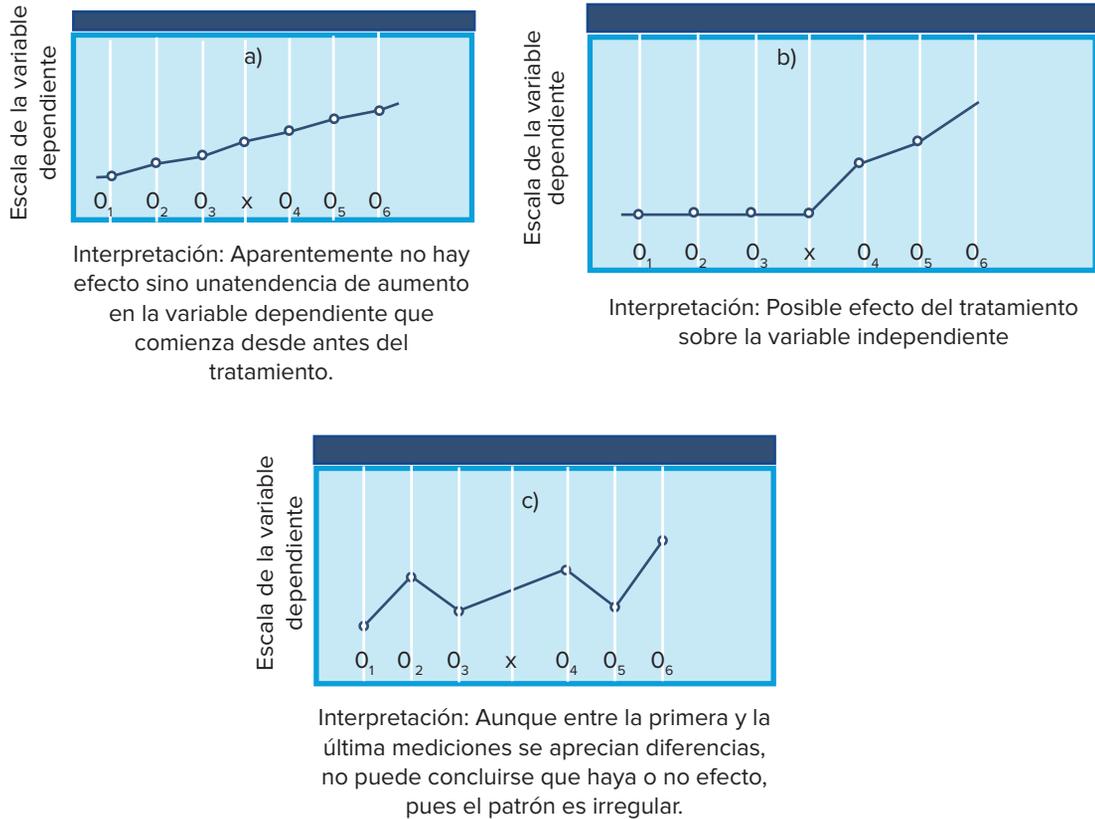
Las series cronológicas de un solo grupo llegan a producir diversos patrones de resultados. A manera de ejemplo podríamos tener los patrones que se muestran en la **figura 5.19** (algunos de los cuales fueron expuestos en las series cronológicas experimentales).

En las series cronológicas de un único grupo debe tomarse muy en cuenta que no se tiene punto de comparación (grupo de control); por lo tanto, la interpretación del patrón de la variable dependiente (o patrones de las variables dependientes) tiene que ser muy cuidadosa, y debe analizarse si no han actuado o interactuado otras posibles causas, además del tratamiento experimental o variable independiente. La historia y el hecho de que el grupo sea atípico son riesgos que se afrontan cuando se emplea este diseño, al igual que la instrumentación. Normalmente, este diseño cuasiexperimental se utiliza con propósitos correlacionales y no explicativos.

Series cronológicas cuasiexperimentales con múltiples grupos

Estos diseños pueden adoptar la estructura de las series cronológicas experimentales, con la diferencia de que en estas últimas la asignación de los individuos a los grupos es aleatoria, mientras que en las cuasiexperimentales tenemos grupos intactos. Por lo tanto, ocurrirían las mismas variaciones⁵ que se muestran en la **figura 5.20**.

⁵ Recomendamos al lector que revise los apartados relativos a las series cronológicas experimentales antes de leer este apartado. En ellos podrá notar que los diseños son los mismos, salvo que en los esquemas y diagramas de las series experimentales aparece el símbolo "R" de asignación al azar. Por lo tanto, aquí se omiten explicaciones, interpretaciones y ejemplos; de lo contrario, pecaríamos de redundantes. Solo nos limitaremos a esquematizar los diseños, sin explicaciones, porque lo que puede decirse es lo mismo que lo que dijimos en las series cronológicas experimentales. Desde luego, debe resaltarse que en las series cuasiexperimentales los grupos son intactos y es necesario asegurar que los grupos sean equiparables.



● **Figura 5.19** Ejemplos de resultados en series cronológicas de un solo grupo.

Sin prepruebas y grupo de control

G ₁	X ₁	01	02	03
G ₂	X ₂	04	05	06
G ₃	X ₃	07	08	09
G ₄	—	010	011	012

G ₁	O ₁	O ₂	X ₁	O ₃	O ₄	X ₁	O ₅	O ₆	O ₇	X ₁	O ₈	O ₉	O ₁₀
G ₂	O ₁₁	O ₁₂	—	O ₁₃	O ₁₄	—	O ₁₅	O ₁₆	O ₁₇	—	O ₁₈	O ₁₉	O ₂₀

Con prepruebas y grupo de control

G ₁	O ₁	X ₁	O ₂	X ₁	O ₃	X ₁	O ₄	X ₁	O ₅	O ₆
G ₂	O ₇	X ₂	O ₈	X ₂	O ₉	X ₂	O ₁₀	X ₂	O ₁₁	O ₁₂
G ₃	O ₁₃	—	O ₁₄	—	O ₁₅	—	O ₁₆	—	O ₁₇	O ₁₈

● **Figura 5.20** Ejemplos de diseños cuasiexperimentales con series cronológicas.

Series cronológicas cuasiexperimentales con repetición de estímulo

Estas series también son similares a sus correspondientes experimentales, pero con grupos intactos. Así, tendríamos los siguientes diagramas para ilustrarlas:

Series cronológicas cuasiexperimentales con tratamientos múltiples

Al igual que en los casos anteriores, estas series son similares a sus correspondientes experimentales, solo que con grupos intactos. Por lo tanto, tendríamos diagramas como estos:

G ₁	X ₁	O ₁	O ₂	X ₂	O ₃	O ₄	X ₃	O ₅	O ₆	O ₇
G ₂	X ₂	O ₈	O ₉	X ₁	O ₁₀	O ₁₁	X ₃	O ₁₂	O ₁₃	O ₁₄
G ₃	X ₃	O ₁₅	O ₁₆	X ₂	O ₁₇	O ₁₈	X ₁	O ₁₉	O ₂₀	O ₂₁
G ₄	X ₂	O ₂₂	O ₂₃	X ₃	O ₂₄	O ₂₅	X ₁	O ₂₆	O ₂₇	O ₂₈
G ₅	X ₁	O ₂₉	O ₃₀	X ₃	O ₃₁	O ₃₂	X ₂	O ₃₃	O ₃₄	O ₃₅
G ₆	X ₃	O ₃₆	O ₃₇	X ₁	O ₃₈	O ₃₉	X ₂	O ₄₀	O ₄₁	O ₄₂

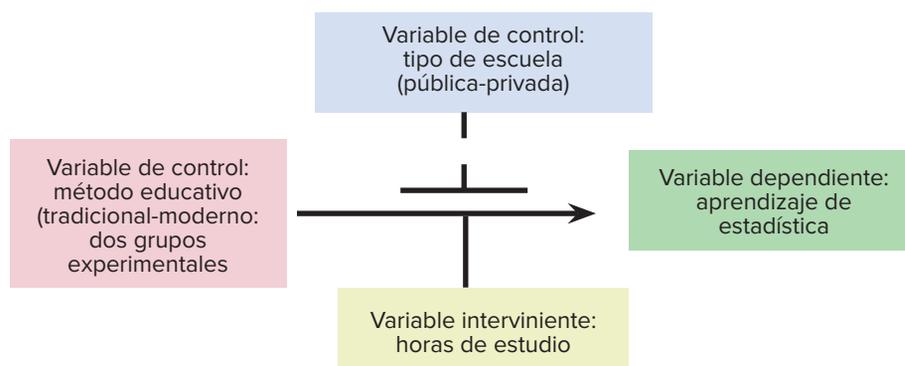
Tipos de variables en experimentos y cuasiexperimentos

Como complemento de lo que se menciona en el libro sobre variables y experimentos, solo queremos ahondar un poco sobre los tipos de variables en los diseños experimentales y cuasiexperimentales. Los principales tipos de variables son:

1. Independiente: tratamiento experimental que provoca efectos (causa). Resulta la variable manipulada.
2. Dependiente: efecto o consecuencia (provocado/a por el tratamiento o variable independiente). Es la variable medida.
3. Interviniente: moderador de la relación causal entre la variable independiente y dependiente. Si no se conoce su efecto o no se controla, el experimento puede invalidarse.
4. Explicaciones rivales o fuentes de invalidación interna (pueden ser variables independientes o intervinientes): su influencia debe conocerse o controlarse pues, de no ser así, el experimento puede invalidarse. Asimismo, es factible que se combine con otras variables para afectar a la dependiente.
5. Variable de control: influye en la dependiente, pero es neutralizada por el diseño o por los procedimientos estadísticos.

La diferencia entre la variable de control e interviniente reside en que en la primera se neutralizan sus efectos, y en la segunda estos se conocen (Creswell, 2013a y Clark y Shadish, 2007). Un ejemplo sería el que se muestra en la **figura 5.21**.

La variable independiente se manipula (a un grupo se le expone a un método, al otro a un método distinto). El tipo de escuela se controla asignando a los dos grupos (en la misma proporción) alumnos de escuelas públicas y privadas por igual (la composición de



● **Figura 5.21** Diferencia entre variables interviniente y de control.

cada uno sería: 50% de estudiantes de instituciones públicas y 50% de escuelas privadas). Los efectos de la variable interviniente se conocen cuando se mide el número de horas dedicadas al estudio (con su introducción al análisis). La variable dependiente se mide. Los análisis estadísticos ayudan a esclarecer las relaciones entre todas las variables.

Otros ejemplos podrían ser:

- Variable independiente: tipo de maquinaria (A y B).
- Variable dependiente: productividad (número de piezas producidas en un cierto periodo).
- Variable de control: capacitación del equipo de trabajo (número de horas y cursos).
- Variable interviniente: experiencia del equipo de trabajo (años).
- Variable independiente: tipo de tratamiento médico (A y B).
- Variable dependiente: presión arterial.
- Variable de control: género.
- Variable interviniente: nivel de ejercitación (horas y tipo).

Nota final: Los pasos de un diseño cuasiexperimental son los mismos que los de uno experimental.

Diseños no experimentales causales

En el capítulo 7 de Metodología de la investigación se señaló que, en ocasiones, en los diseños no experimentales causales se reconstruyen las relaciones a partir de la(s) variable(s) dependiente(s), en otras a partir de la(s) independiente(s) y en otras más sobre la base de variabilidad amplia de las independientes y dependientes (Maggetti, Gilardi y Radaelli, 2013 y León y Montero, 2003). Al primer caso se le conoce como retrospectivo, al segundo como prospectivo y al tercero como causalidad múltiple. Veámoslos más a detalle.

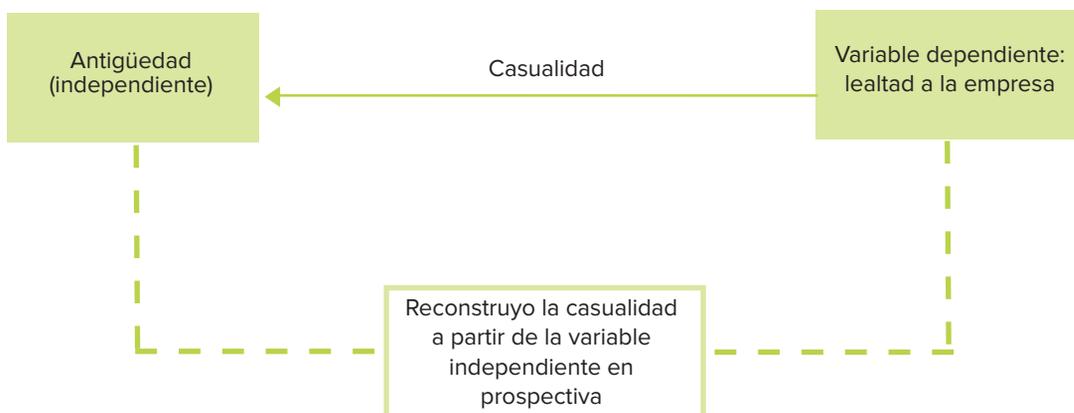
Spongamos que mi interés es analizar las causas por las cuales algunos clientes, y otros no, han utilizado el crédito que les fue otorgado por una cadena de tiendas departamentales. En este caso, la variable dependiente tiene dos niveles: a) clientes que sí han utilizado su crédito y b) clientes que no lo han hecho. Empleamos la base de datos de los clientes y los agrupamos en el nivel que les corresponde. Luego, les preguntamos a

quienes sí han empleado el crédito los motivos que tuvieron para ello; del mismo modo, a quienes no lo han hecho, les preguntamos las razones por las que no lo han utilizado. Así determinamos las causas que nos importan. El estudio podría diagramarse tal como se muestra en la **figura 5.22**. El estudio causal se desarrolla en un momento particular y único.



● **Figura 5.22** Ejemplo de una reconstrucción causal prospectiva.

Veamos ahora una investigación causal prospectiva: imaginemos que deseamos indagar si la variable antigüedad provoca o no mayor lealtad a la empresa y por qué. Como primer paso, dividimos a los empleados con relación a la variable independiente: *a*) muy alta antigüedad (25 o más años en la organización), *b*) alta antigüedad (16 a 24 años), *c*) mediana antigüedad (9 a 15 años), *d*) baja antigüedad (cuatro a ocho años), *e*) muy baja antigüedad (uno a tres años) y *f*) ingreso reciente (un año o menos). Posteriormente, medimos los niveles de lealtad y cuestionamos a los empleados sobre cómo la antigüedad ha generado o no mayor lealtad. De esta forma determinamos los efectos de interés (**figura 5.23**).



● **Figura 5.23** Ejemplo de una reconstrucción causal retrospectiva.



En los diseños donde se reconstruyen las relaciones sobre la base de variabilidad amplia de las independientes y dependientes no se parte de una variable en especial ni de grupos, sino que se evalúa la estructura causal completa (las relaciones en su conjunto. Vea la **figura 5.24**).



● **Figura 5.24** Modelo mediatizador del clima organizacional.

Todos los estudios transeccionales causales nos brindan la oportunidad de predecir el comportamiento de una o más variables a partir de otras, una vez que se establece la causalidad. A estas últimas se les denomina variables predictoras. Tales diseños requieren de análisis multivariados, que se mencionan en el capítulo 8 de este centro de recursos: Análisis estadístico: segunda parte. En la figura 5.19 simplemente incluimos un ejemplo de una estructura causal compleja. Lo importante es que se comprenda cómo en ocasiones se analizan múltiples variables y secuencias causales.

En el caso del modelo de la figura 5.24, las percepciones sobre las variables o dimensiones del clima organizacional (trabajo, papel que se desempeña, líder o superior, grupo de trabajo y elementos de la organización) influyen en la motivación y el desempeño, pero con la mediación de las actitudes hacia el trabajo (satisfacción en el trabajo, involucramiento en el trabajo y el compromiso con la empresa o institución). Es decir, hay dos niveles de variables intervinientes: las del clima y las actitudes hacia el trabajo. El modelo se fundamenta en Parker et al. (2003) y Hernández-Sampieri (2005). Las percepciones psicológicas del clima son las variables predictoras iniciales.