

# **DISEÑOS CUASI-EXPERIMENTALES Y LONGITUDINALES**

Roser Bono Cabré

Departamento de Metodología de les Ciencias del Comportamiento

Facultad de Psicología

Universidad de Barcelona

# ÍNDICE

<b>1. Metodología cuasi-experimental.....</b>	<b>2</b>
1.1. Concepto de cuasi-experimento.....	2
1.2. Características básicas.....	4
1.3. Estrategias cuasi-experimentales.....	13
1.4. Clasificación general.....	15
<b>2. Diseños cuasi-experimentales.....</b>	<b>19</b>
2.1. Diseños de grupo de control no equivalente (DGNE).....	20
2.2. Diseños de discontinuidad en la regresión (DDR).....	34
<b>3. Diseños longitudinales.....</b>	<b>40</b>
3.1. Diseños de series temporales interrumpidas (DSTI).....	40
3.2. Diseños de medidas repetidas (DMR).....	46
3.3. Diseños de cohortes (DC).....	54
3.4. Diseños en panel (DP).....	62
<b>4. Modelado estadístico.....</b>	<b>66</b>
4.1. Modelos estadísticos.....	66
4.2. Fases del modelado estadístico.....	68
4.3. Inferencia estadística: contraste de hipótesis.....	71
<b>5. Referencias bibliográficas.....</b>	<b>76</b>

# 1. METODOLOGÍA CUASI-EXPERIMENTAL

Si bien se ha mejorado metodológica y estadísticamente la posibilidad de tener en cuenta y controlar las fuentes de confundido extrañas, el enfoque cuasi-experimental es mucho más vulnerable a las amenazas contra la validez inferencial, en comparación con la estrategia experimental. Además, en un cuasi-experimento pueden plantearse más hipótesis alternativas que se ajusten a los datos. No obstante, por otro lado, existe una mayor generalización de los resultados a otras situaciones distintas a la de investigación.

A pesar de las limitaciones que presenta la metodología cuasi-experimental, la correcta planificación y ejecución de un cuasi-experimento, que en muchas situaciones constituye la única alternativa de trabajo, puede aportar una información adecuada sobre el impacto de un tratamiento o evolución del cambio.

## 1.1. Concepto de cuasi-experimento

Los diseños cuasi-experimentales, principales instrumentos de trabajo dentro del ámbito aplicado, son esquemas de investigación no aleatorios. Dado la no aleatorización, no es posible establecer de forma exacta la equivalencia inicial de los grupos, como ocurre en los diseños experimentales. Cook y Campbell (1986) consideran los cuasi-experimentos como una alternativa a los experimentos de asignación aleatoria, en aquellas situaciones sociales donde se carece de pleno control experimental:

Los cuasi-experimentos son como experimentos de asignación aleatoria en todos los aspectos, excepto en que no se puede presumir que los diversos grupos de tratamiento sean inicialmente equivalentes dentro de los límites del error muestral (p. 142).

Tal como afirma Campbell (1988), "podemos distinguir los cuasi-experimentos de los experimentos verdaderos por la ausencia de asignación aleatoria de las unidades a los tratamientos" (p. 191).

Una definición que incluye las características más relevantes de la

metodología cuasi-experimental es la ofrecida por Pedhazur y Schmelkin (1991):

¿Qué es un cuasi-experimento? Es una investigación que posee todos los elementos de un experimento, excepto que los sujetos no se asignan aleatoriamente a los grupos. En ausencia de aleatorización, el investigador se enfrenta con la tarea de identificar y separar los efectos de los tratamientos del resto de factores que afectan a la variable dependiente (p. 277).

La descripción de diseño cuasi-experimental propuesta por Hedrick *et al.* (1993) es la siguiente:

Los diseños cuasi-experimentales tienen el mismo propósito que los estudios experimentales: probar la existencia de una relación causal entre dos o más variables. Cuando la asignación aleatoria es imposible, los cuasi-experimentos (semejantes a los experimentos) permiten estimar los impactos del tratamiento o programa, dependiendo de si llega a establecer una base de comparación apropiada (p. 58).

En la misma línea que los autores precedentes, Arnau (1995c) define la perspectiva cuasi-experimental de la siguiente forma:

Aquellos sistemas de investigación donde el criterio de asignación de los sujetos o unidades a las condiciones de tratamiento o condiciones de estudio no se rige por las leyes del azar (p. 15).

Del mismo modo, Ato (1995b) considera a la metodología cuasi-experimental como:

Un conjunto de técnicas de diseño y análisis estadístico para afrontar situaciones donde no es posible o no es ético aplicar la metodología experimental, o donde los estrictos requisitos del método experimental no se satisfacen (p. 45).

En las definiciones anteriores se observa que la característica principal

de las investigaciones cuasi-experimentales es la ausencia de aleatorización de los tratamientos y, por lo tanto, la carencia de un control total sobre la situación. Al interpretar los resultados de un cuasi-experimento, hay que considerar la posibilidad de que se deban a otros factores no tenidos en cuenta (Cook y Campbell, 1986). En efecto, en un diseño cuasi-experimental, en comparación con los experimentales, hay más hipótesis alternativas que pueden ajustarse a los datos. Por ello, es imprescindible que el investigador tenga, en la medida de lo posible, un conocimiento de las variables específicas que el diseño cuasi-experimental utilizado no sea capaz de controlar.

Debido al deficiente control que implican las investigaciones cuasi-experimentales, algunos autores, como por ejemplo Van Dalen y Meyer (1971), se refieren a ellas con el nombre de investigaciones con *control parcial*. Otros metodólogos prefieren utilizar el término de *estudios no aleatorizados* (Anderson *et al.*, 1980).

## **1.2. Características básicas**

En términos generales, los diseños cuasi-experimentales presentan una serie de carencias y dificultades, lo que conlleva posibles fuentes de invalidez. Por esta razón, Campbell y Stanley (1963, 1966) recomendaron utilizarlos únicamente cuando no se puedan asignar aleatoriamente los sujetos a las diferentes condiciones.

Antes de especificar las características propias de la cuasi-experimentación, es necesario recordar que en todo trabajo de investigación debe evaluarse lo que, de forma genérica, es conocido por validez. Como destacan Judd y Kenny (1981), el propósito de toda investigación social es conseguir una información exacta y generalizable, lo cual equivale a afirmar que sea válida. Cook y Campbell (1979) aportan una definición del término validez con la siguiente expresión: "usaremos los conceptos de *validez* e *invalidez* para referirnos a la mejor aproximación disponible a la verdad o falsedad de las proposiciones" (p. 37). Nuestro objetivo no es entrar en la polémica suscitada a raíz de las diferentes tipologías sobre validez (véase Mark, 1986), tan sólo recordar lo que se entiende por validez interna y externa.

La *validez interna* se refiere a la posibilidad de poder derivar conclusiones consistentes acerca de la efectividad de un tratamiento o programa social, y la *validez externa* hace referencia al alcance y extensión que tienen los resultados. La mayoría de investigaciones dentro del ámbito aplicado ponen el énfasis en la validez externa. Sin embargo, no pueden tomarse decisiones acerca de la validez externa sin tener en cuenta, en la valoración de los resultados, la validez interna. En otras palabras, es necesario hallar un equilibrio o compromiso entre estos dos tipos de validez. Quizá, como regla práctica, puede aconsejarse que cuando el tratamiento es innovador lo importante es destacar la validez interna, y cuando la intervención ha demostrado ser efectiva, en alguno de los grupos, lo interesante sería resaltar la validez externa. Para un estudio detallado de la validez puede consultarse a Ato (1991, 1995b), Ato y Rabadán (1991), Cook y Campbell (1976, 1979) y Cook *et al.* (1990).

Pedhazur y Schmelkin (1991) señalan que sin control no hay validez y que, por lo tanto, estos dos términos constituyen los aspectos centrales de un diseño de investigación. De ahí, la importancia que adquieren las técnicas de control dentro del marco de la investigación científica. No obstante, cuando se trata de reflejar lo que sucede en el ambiente natural, se está obligado a trabajar en condiciones en las que es difícil el control experimental. Este es el caso típico de lo que se conoce por investigación aplicada.

El objetivo primordial de toda investigación científica es la inferencia de relaciones causales entre las variables. Por lo tanto, la evaluación del efecto de la variable independiente es la meta a conseguir tanto por la experimentación como por la cuasi-experimentación. La correcta evaluación del tratamiento depende de si es posible la asignación aleatoria de las unidades de observación a los valores que toma la variable independiente, y del control de los posibles factores extraños productores de cambio en los registros de la variable de medida (Campbell y Boruch, 1975; Gilbert, Light y Mosteller, 1975; Riecken y Boruch, 1974). Sin embargo, en muchos contextos sociales y comportamentales, a causa de una gran variedad de problemas prácticos, éticos y políticos, el supuesto de aleatorización no puede cumplirse (Cohen, 1975; Reynolds y Temple, 1995; Rossi y Freeman, 1989; Suchman, 1967; Weiss,

1972b) y, además, es difícil conseguir una completa especificación de los factores causales para que puedan ser aislados o controlados mediante técnicas experimentales (Arnau, 1995c). De acuerdo con esta caracterización, existen diferencias entre dos grandes estrategias de investigación: la experimental y la cuasi-experimental. Dichas estrategias no poseen la misma validez inferencial o, en otros términos, no tiene igual fuerza probatoria. Así, de este modo, en los diseños experimentales, cualquier diferencia entre grupos puede atribuirse a la acción del tratamiento, dado que la aleatorización los ha igualado en todos los aspectos (excepto en la variable independiente). Por el contrario, en la investigación cuasi-experimental, debido a la ausencia de aleatoriedad, no queda garantizada la exclusión de factores extraños capaces de contaminar los resultados. Únicamente cuando el grupo control es similar al grupo de tratamiento en todos los aspectos substanciales, la investigación cuasi-experimental guarda una gran semejanza con los experimentos aleatorizados (Rossi y Freeman, 1989).

Según Kenny (1975a, 1979), el no poder utilizar la aleatoriedad puede deberse a que: a) el tratamiento sea administrado a un grupo (aula, colegio o sistema escolar), tomándose otro grupo como control; b) se había planificado realizar un experimento verdadero, pero debido a la mortalidad, contaminación de las unidades del grupo control por artefactos experimentales o por variaciones en el tratamiento, el experimento verdadero se ha convertido en un cuasi-experimento; c) por escasez de recursos sólo se aplica el tratamiento a un grupo pre-seleccionado, y d) los sujetos autoseleccionan su propio nivel de tratamiento. En cualquiera de estos casos pueden darse diferencias entre las puntuaciones pre-tratamiento y, por lo tanto, los grupos no son comparables. Reichardt (1979) afirma que la no equivalencia implica que en el supuesto de llevarse a cabo varias veces la misma selección no aleatoria, los grupos diferirían en una gran variedad de formas.

El hecho de que la validez interna de los cuasi-experimentos sea menor que la de los experimentos verdaderos, no es un argumento en contra de la utilización de los diseños cuasi-experimentales (Kenny, 1975a). La diferencia entre experimentación y cuasi-experimentación se muestra en la figura 1.1 (para el caso de un diseño con una variable independiente y otra dependiente).

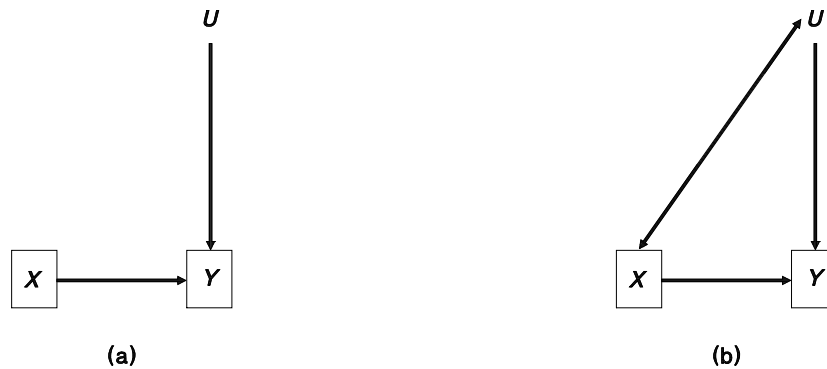


Figura 1.1. Representación gráfica de a) una investigación experimental y b) cuasi-experimental (modificada de Pedhazur y Schmelkin, 1991; p. 280).

La figura 1.1a representa un experimento donde el tratamiento (simbolizado por  $X$ ) afecta a la variable dependiente (simbolizada por  $Y$ ). En el supuesto de asignación aleatoria de los sujetos o unidades a los diferentes niveles de  $X$ , se asume que los grupos son equivalentes en todas las otras variables (incluidas en  $U$ ) capaces de afectar a  $Y$ . En este caso, no se espera correlación entre  $X$  y  $U$ .

Cuando el modelo de la figura 1.1a no es verificado empíricamente, entonces se adoptan modelos alternativos para una adecuada explicación de los datos. En la figura 1.1b se muestra esquemáticamente un cuasi-experimento. A diferencia del caso anterior, la asignación de los sujetos a las distintas categorías de  $X$  no se realiza aleatoriamente y, por lo tanto, los grupos no son equivalentes, con lo que pueden aparecer errores de especificación por la omisión de variables relevantes (incluidas en  $U$ ) correlacionadas con  $X$ . Tales correlaciones pueden ser debidas a que  $X$  afecte a la variable omitida o viceversa, o que ambas estén afectadas por otra variable. Este hecho lleva a estimaciones sesgadas de los efectos de  $X$ . Para evitar este tipo de sesgos es necesario explicar el modelo teórico de las relaciones entre las variables, lo cual tiene importantes implicaciones en relación a los intentos de ajuste de las diferencias iniciales entre grupos que reciben distintos



tratamientos.

Queda claro que el inconveniente principal de los diseños cuasi-experimentales es la ausencia o insuficiencia de control en la asignación de los sujetos a los grupos. Esta falta de control puede deberse a varios motivos. Pedhazur y Schmelkin (1991) destacan las siguientes razones: que el investigador aparezca cuando el estudio ya ha empezado o bien cuando ha terminado, restricciones institucionales, presiones políticas, consideraciones éticas y factores económicos.

La asignación no aleatoria puede llevar no sólo a conclusiones erróneas sino también a ir en contra del sentido común (Campbell, 1988). Campbell y Boruch (1975) ponen un ejemplo que refleja claramente esta limitación de la investigación cuasi-experimental:

Si un estudio cuasi-experimental muestra que la gente que ha recibido psicoterapia tiene una tasa de suicidio mayor que la de un grupo control, es lógico tener dudas para aceptar la conclusión de que la psicoterapia causa suicidios. Está claro que ninguna comparación puede corregir correctamente el hecho de que los sujetos que solicitan psicoterapia tienen mayor tendencia al suicidio (p. 203).

En la mayoría de circunstancias, la participación en un cuasi-experimento y la recepción de tratamientos específicos viene determinada por un proceso de autoselección o voluntarismo de los sujetos. Por ejemplo, en el caso presentado por Campbell y Boruch (1975), los pacientes que solicitan una terapia suelen ser los más perturbados. Lo mismo ocurre en la asignación no aleatoria. Así, los sujetos que forman parte de grupos naturales ya construidos o intactos (hospitales, distritos, escuelas, clases, etc.) se asignan a un determinado tratamiento porque en base a su nivel en la variable dependiente (por ejemplo, logros académicos, peso, alcoholismo, etc.) se considera que tienen mayores probabilidades de beneficiarse del tratamiento, es decir, que responderán de acuerdo con los objetivos de la investigación. Tanto la autoselección como la asignación no aleatoria pueden sesgar los efectos del tratamiento.

A partir de lo expuesto, lo ideal sería comparar el grupo sometido a tratamiento con un grupo control equivalente. Esto sólo se consigue con la asignación aleatoria de los sujetos a los grupos de control y tratamiento. De este modo, tendríamos un control físico o aleatorizado de las variables extrañas. En ausencia de aleatoriedad, nos hallamos con comparaciones entre grupos no equivalentes. Aunque la asignación no aleatoria o la autoselección no lleve a diferencias en la variable dependiente respecto a la situación anterior al tratamiento, los grupos pueden diferenciarse en distintos factores desconocidos (variables de selección o de sesgo) y, en muchos casos, imposibles de conocer que afectan a la variable dependiente o a la susceptibilidad del sujeto para beneficiarse de un determinado tratamiento. En otras palabras, puede existir una interacción entre selección y tratamiento. La omisión de variables relevantes correlacionadas con la variable independiente constituye lo que Pedhazur y Schmelkin (1991) han denominado *error de especificación*, lo cual genera estimaciones sesgadas de los efectos de la variable de tratamiento.

Los diseños cuasi-experimentales carecen de potencia suficiente para el rechazo de explicaciones causales rivales. Arnau (1995c) y Dwyer (1983) proponen, dentro de la estrategia transversal, tres modelos alternativos capaces de explicar la relación entre la variable X y la variable Y: a) modelo de espureidad (figura 1.2); b) modelo de causación inversa (figura 1.3), y c) modelo de causación mediatizada (figura 1.4).

De acuerdo con la figura 1.2 (modelo de espureidad), la correlación entre X e Y queda anulada por una tercera variable (simbolizada por Z). Conforme al modelo de causación inversa (figura 1.3), la dirección de la causalidad va de Y a X (un cambio en Y ocasiona un cambio en X). En este caso, las investigaciones cuasi-experimentales no pueden recoger, de ningún modo, la direccionalidad de los efectos causales. Por último, en el modelo de causación mediatizada (figura 1.3) los cambios originados en X influyen en Y solamente cuando tales cambios se dan a través de una variable mediadora o interviniente Z.

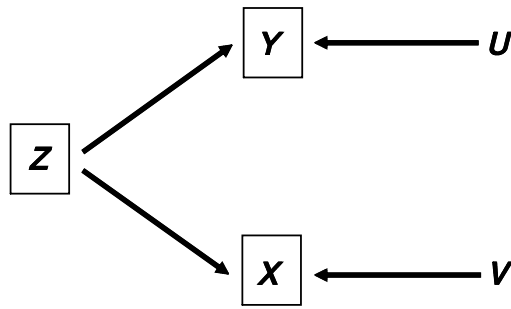


Figura 1.2. Modelo de espureidad o de una tercera variable, donde  $U$  y  $V$  son factores de perturbación.

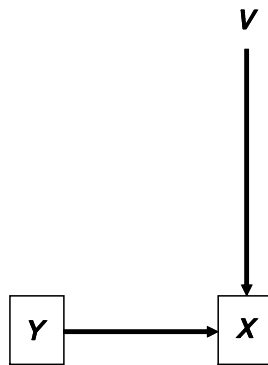


Figura 1.3. Modelo de causación inversa.

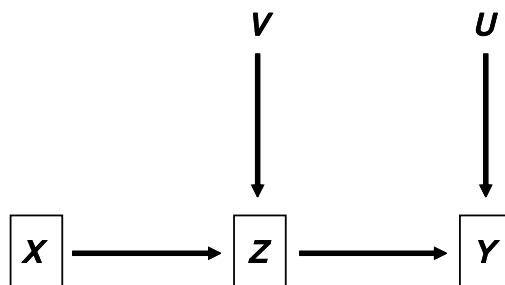


Figura 1.4. Modelo de causación mediatizada.

De los modelos de explicación alternativos, se concluye que los diseños cuasi-experimentales de carácter transversal no garantizan la validez interna, y tiene pocos indicios de causalidad. En otras palabras, varios modelos alternativos son capaces de generar los mismos datos y, por lo tanto, es difícil establecer relaciones de causalidad entre las variables. En cambio, la estrategia longitudinal permite rechazar con más fuerza algunos de los modelos de explicación alternativos asociados a la situación transversal (Bentler, 1978). Como ha señalado Arnau (1995c):

Cuando en el diseño se introduce la dimensión tiempo se abren nuevas perspectivas explicativas, puesto que se puede aprovechar la información obtenida del sentido del cambio que presentan los datos u observaciones a lo largo de esta dimensión (p. 27).

En este sentido, dentro del ámbito cuasi-experimental, los diseños longitudinales constituyen los instrumentos más potentes aplicados a las ciencias sociales y del comportamiento.

Dwyer (1983) destaca tres ventajas principales de los diseños longitudinales, en comparación con los transversales. La primera hace referencia a que los efectos causales son unidireccionales: un cambio en la variable Y, en un punto posterior del tiempo, no puede afectar a la variable X, en un punto del tiempo anterior. En segundo lugar, existen disposiciones longitudinales -tales como los diseños en panel- que permiten probar los modelos de una tercera variable o de espureidad, sin necesidad de identificar y medir esta tercera variable. Finalmente, es posible medir los cambios de cada sujeto o caso, sin inferirlos a partir de la diferencia entre los individuos.

Los diseños longitudinales brindan más posibilidades que los transversales en cuanto a la inferencia de relaciones causales. No obstante, no siempre la garantía es completa (Arnau, 1995c; Rutter, 1988). Aunque los diseños transversales tengan una deficiente potencia inferencial, no por ello debe renunciarse a su aplicación. Una de las principales tareas del investigador es identificar cuáles son los factores de sesgo responsables de la

desigualdad entre grupos y, posteriormente, corregir o ajustar los datos a los valores de las variables de sesgo mediante técnicas estadísticas (Arnau, 1995a; Reynolds y Temple, 1995). Algunos procedimientos para conseguir comparaciones válidas son más exitosos que otros. Quizá el ACOVAR es el más utilizado con objeto de proporcionar un estimador insesgado del efecto del tratamiento (Weisberg, 1979). Sin embargo, como destaca Lord (1967):

...no existe ningún procedimiento lógico o estadístico con el que se puedan tener en cuenta de forma correcta las diferencias pre-existentes no controladas entre grupos (p. 305).

Algunos autores, como por ejemplo Campbell y Boruch (1975), interpretan la afirmación anterior como que los ajustes estadísticos no son del todo fiables y, en consecuencia, es necesario utilizar más experimentos aleatorios. Tal postura es poco realista, sobre todo con intervenciones sociales a gran escala donde la aleatoriedad no es nada práctica (Weisberg, 1979).

Hasta aquí sólo hemos hecho mención de las dificultades que presenta la investigación cuasi-experimental. Por otra parte, los diseños cuasi-experimentales poseen un alto nivel de disponibilidad y flexibilidad, puesto que su aplicación es aconsejable cuando, debido a las limitaciones del contexto, no se puede utilizar un diseño experimental. Por tanto, en aquellos casos donde no es posible aplicar el ideal de experimentación, no por ello deberá renunciarse a investigar e intentar utilizar diseños que, aunque carezcan de un completo control, pueden ofrecer la posibilidad de obtener resultados científicamente válidos.

De lo expuesto en este apartado, podemos establecer de forma sintética las características básicas de los diseños de investigación aplicados, en función de sus objetivos, efectos inferidos, factores extraños, selección de las unidades, validez enfatizada y alcance de los resultados (tabla 1.1).

Tabla 1.1. Características de los diseños cuasi-experimentales.

<b>OBJETIVOS</b>	Estudio del impacto de los tratamientos y de los procesos de cambio intra e inter-individuales.
<b>EFFECTOS INFERIDOS</b>	Mayor riesgo de espureidad en los diseños transversales
<b>FACTORES EXTRAÑOS</b>	Escaso control
<b>SELECCIÓN DE LAS UNIDADES</b>	Sesgada
<b>VALIDEZ ENFATIZADA</b>	Validez externa
<b>ALCANCE DE LOS RESULTADOS</b>	Muy generalizables

### 1.3. Estrategias cuasi-experimentales

Los diseños que mejor se ajustan a la investigación aplicada son los cuasi-experimentales, por lo que, a continuación, exponemos las estrategias que pueden adoptar este tipo de diseños.

Dentro de la perspectiva cuasi-experimental, los diseños pueden dividirse según dos estrategias de recogida de datos. La primera corresponde a la estrategia *transversal* o *entre-sujetos*, basada en la comparación de grupos no equivalentes. La segunda estrategia, la *longitudinal*, consiste en llevar a cabo comparaciones de tipo *intra-sujeto*, es decir, registrar la misma respuesta a lo largo de una serie de puntos en el tiempo. La diferencia entre ambas estrategias está en que la transversal se basa en el sentido estático de la comparación, y la longitudinal se caracteriza por su naturaleza dinámica o, en otros términos, enfatiza el carácter temporal de la comparación.

### **1.3.1. Estrategia transversal**

En los *diseños transversales*, la conducta de los sujetos (o unidades de observación) se registra en un punto de corte en el tiempo. De ahí el porqué, a veces, estos diseños son conocidos con el nombre de *diseñostrans-seccionales* (Dwyer, 1983). Dado que, por otra parte, los sujetos no se asignan aleatoriamente a los diferentes grupos de tratamiento, ni la variable independiente es, por lo general, activamente manipulada, los registros obtenidos tienen un carácter pasivo.

Dentro de la estrategia transversal, las unidades pueden ser seleccionadas siguiendo a procedimientos conocidos o desconocidos. En cualquiera de los casos, nunca interviene el azar, por lo que los grupos son inicialmente no equivalentes. La forma más importante de diseños caracterizados porque la variable de asignación no es aleatoria ni conocida son los de grupo control no equivalente. Cuando la variable de asignación es conocida, se tiene como principal estructura de investigación el diseño de discontinuidad en la regresión.

### **1.3.2. Estrategia longitudinal**

Por lo que respecta a los *diseños longitudinales*, éstos se dividen, según las unidades de observación, en diseños de un solo sujeto o unidad observacional y diseños de un grupo o  $k$  grupos de sujetos. Cuando se registra el comportamiento de un sujeto, cabe destacar que se toma una cantidad considerable de medidas u observaciones de forma sucesiva y secuencial a lo largo de un amplio período de tiempo (como por ejemplo en los diseños de series temporales). Por el contrario, en los diseños de un grupo o varios grupos de sujetos, los registros se recogen a partir de una serie de ocasiones de observación fijas y espaciadas en el tiempo (por lo general, dos o más puntos de observación). Dentro de este segundo subgrupo, tendríamos diseños tales como los de medidas repetidas, los diseños de cohortes y los diseños en panel. Estos últimos requieren como mínimo el registro simultáneo de dos variables y de un mismo grupo de sujetos en cortes discretos del tiempo.

En resumen, los diseños cuasi-experimentales de carácter transversal no utilizan en la formación de los grupos el criterio de azar. Debido a ello, estos diseños suelen estar sometidos, entre otros factores, a los efectos del sesgo de selección. Por otra parte, la estrategia longitudinal, que incorpora la variable tiempo, se caracteriza por la cantidad y amplitud del intervalo de registro. El estudio de ambas estrategias o procedimientos de investigación -transversal y longitudinal- permitirá conocer la gran flexibilidad que posee la cuasi-experimentación, así como las distintas estructuras de estos diseños considerados como una alternativa a la investigación experimental, dentro del contexto social y comportamental aplicado.

#### **1.4. Clasificación general**

Siguiendo a Judd y Kenny (1981), tres son los aspectos esenciales a tener en cuenta en la clasificación de los diseños cuasi-experimentales: a) la estrategia de comparación y obtención de datos (transversal o longitudinal); b) el procedimiento de asignación de las unidades a los tratamientos (desconocido o conocido), y c) la existencia o no de medidas pre-test. Veamos a continuación cada uno de estos factores con más detalle:

- 1) El primer factor se refiere al modo en que se realizan las comparaciones con objeto de inferir el efecto de la variable de tratamiento. Se pueden dar dos estrategias: observar diferentes sujetos en las condiciones de tratamiento y control, o bien exponer los sujetos a todas las condiciones de estudio. En el primer caso, el efecto del tratamiento se estiman a través de comparaciones entre-sujetos y en el segundo caso, la inferencia del efecto del tratamiento se lleva a cabo mediante comparaciones intra-sujetos.
- 2) Un segundo factor es la forma de asignación de las unidades a las condiciones de tratamiento. En otras palabras, cuál es la regla o variable que ha determinado la formación de los grupos. En los diseños cuasi-experimentales existen dos reglas de asignación



principales: una conocida donde el factor que determina la construcción de los grupos se basa en alguna variable medida (por ejemplo las puntuaciones obtenidas por los sujetos en un pre-test), y una regla de asignación desconocida donde el criterio en función del cual se forman los grupos no es medido ni tampoco se conoce.

- 3) El último factor de diferenciación entre los diseños de investigación cuasi-experimental está relacionado con la presencia o ausencia de medidas pre-tratamiento o pre-test.

Según los factores establecidos por Judd y Kenny (1981), el investigador debe decidir, en primer lugar, qué estrategia de comparación va a utilizar: entre-sujetos o intra-sujetos. Esta primera consideración, nos permite clasificar los diseños cuasi-experimentales en transversales y longitudinales, respectivamente. En segundo lugar, cabe destacar el procedimiento de formación de los grupos (de acuerdo con criterios conocidos o desconocidos). Por último, es factible tener en cuenta el hecho de tomar medidas pre-test. Éstas son indispensables en las comparaciones intra-sujeto y también cuando se utilizan para la formación de los grupos. Muchas veces, las medidas pre-tratamiento permiten constatar la equivalencia inicial de los grupos.

Siguiendo con esta línea, proponemos un modelo general de clasificación basado en la distribución temporal de las observaciones (Arnau, 1997). Así, una o varias mediciones con un intervalo de tiempo muy corto entre ellas genera un diseño transversal. En caso de efectuarse repetidas medidas a lo largo del tiempo nos encontraríamos con un diseño longitudinal. Dentro de cada uno de estos bloques hay distintas variedades de los diseños de investigación aplicados (figura 1.5).

Aunque se han propuesto diversas clasificaciones del diseño cuasi-experimental, todas ellas coinciden en mantener la dicotomía entre dos grandes bloques: los diseños transversales y los diseños longitudinales.

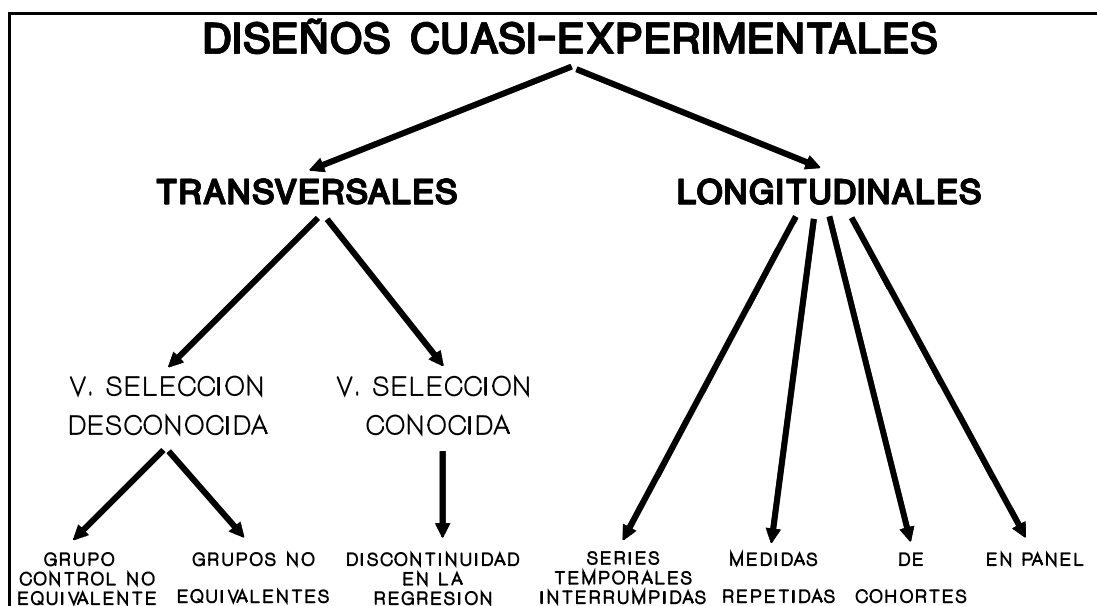


Figura 1.5. Clasificación de los diseños cuasi-experimentales.

En la figura 1.5 se observa que según la variable de selección sea desconocida o conocida, los diseños transversales se dividen en diseños de grupo control no equivalente (DGCNE) o diseños de grupos no equivalentes y diseños de discontinuidad en la regresión (DDR). Los DGCNE tienen la estructura prototípica de un diseño cuasi-experimental. Su formato consta de medidas antes y después de la aplicación del tratamiento. Las medidas antes sirven para ajustar las diferencias iniciales de los dos grupos (control y tratamiento) mediante técnicas estadísticas adecuadas. En los DDR el procedimiento de selección consiste en asignar los sujetos al grupo de tratamiento y control conforme a un punto de corte en la variable pre-tratamiento. Por ejemplo, los sujetos con puntuaciones por debajo del punto de corte son asignados al grupo de control, mientras que los sujetos que se sitúan por encima son asignados al grupo de tratamiento.

Dentro de la estrategia longitudinal, los diseños más utilizados a lo largo de los últimos años son: los diseños de series temporales interrumpidas (DSTI); los diseños de medidas repetidas (DMR), los diseños de cohortes y los diseños en panel.

La característica más relevante de los DSTI es el hecho de que suele trabajarse con unidades individuales (un sólo dato por punto en el tiempo y por variable de medida); aunque también es posible agregar los datos generados por distintas unidades. Los DSTI poseen una estructura donde un conjunto de datos, registrados secuencialmente en el tiempo, es interrumpido por un hecho o intervención. Existen dos modalidades básicas de esta clase de diseños: el DSTI simple y el DSTI de grupo control no equivalente. La primera consiste en la obtención de una gran cantidad de registros seguidos y frecuentes de la variable dependiente en una serie discreta de observaciones, antes y después de la intervención. Una extensión de este diseño es el DSTI de grupo control no equivalente. La ventaja de esta última modalidad es que permite controlar el efecto de cualquier hecho externo que ocurre al mismo tiempo que la intervención, capaz de confundir la acción del tratamiento.

En contraposición a los DSTI, los DMR se caracterizan por la cantidad restringida de datos en intervalos de tiempo más espaciados y la utilización de grupos o muestras de sujetos. Respecto a los diseños de cohortes, éstos permiten estimar los efectos de edad, período y cohorte en un proceso de cambio. Tradicionalmente, el estudio de estos tres factores se ha llevado a cabo mediante la aplicación de diseños transversales, longitudinales y de retardo temporal. El inconveniente principal es que se confunden los efectos de edad, período y cohorte.

Por último, dentro de la estrategia longitudinal, cabe destacar los diseños en panel. La modalidad más utilizada es la de retardos cruzados que se caracteriza por el registro simultáneo de dos variables (de una misma muestra de sujetos) en dos tandas o períodos de tiempo. Esta estructura de diseño tiene por objetivo hallar la dirección de causalidad entre las variables.

## 2. DISEÑOS CUASI-EXPERIMENTALES

De acuerdo con el tipo de estrategia y los objetivos a alcanzar, los diseños cuasi-experimentales se dividen en transversales y longitudinales. Los transversales son equivalentes a los diseños de comparación de grupos, y los longitudinales permiten estudiar los procesos de cambio y sus posibles causas. A su vez, los diseños transversales se categorizan según si la variable que forma los grupos es conocida (diseños de discontinuidad en la regresión) o desconocida (diseños de grupo control no equivalente y diseños de grupos no equivalentes). Por su parte, los diseños longitudinales se clasifican en función de si los intervalos de observación son seguidos y frecuentes (diseños de series temporales), discretos y distantes (diseños de medidas repetidas) o poco frecuentes -dos o tres períodos de observación- y más distantes (diseños en panel).

Las técnicas de análisis que pueden aplicarse a los datos de diseños de investigación aplicados son, fundamentalmente, el análisis de la variancia (AVAR), el análisis de la covariancia (ACOVAR), el análisis de la regresión múltiple (ARM), el análisis de series temporales (modelos ARIMA), el análisis multivariado de la variancia (AMVAR), el análisis de correlaciones cruzadas en panel (ACCP), el análisis de la regresión (AR) y los modelos lineales de ecuaciones estructurales (LISREL). En la tabla 2.1 se relacionan los distintos diseños cuasi-experimentales con las correspondientes técnicas de análisis.

Tabla 2.1. Diseños de investigación aplicados: técnicas de análisis.

DISEÑOS	TÉCNICAS DE ANÁLISIS
De grupo control no equivalente	AVAR, ACOVAR, ARM
De grupos no equivalentes	AVAR, ACOVAR, ARM
De grupos no equivalentes con múltiples covariables	AVAR, ARM, ACOVAR MÚLTIPLE
De discontinuidad en la regresión	AVAR, ACOVAR, ARM
De series temporales interrumpidas	ARIMA
De medidas repetidas	AVAR, AMVAR
De cohortes	AVAR
En panel	ACCP, AR, LISREL

### 2.1. Diseños de grupo control no equivalente (DGCNE)

El DGCNE es una estructura de investigación cuasi-experimental donde existe un grupo de tratamiento y otro de control. En ambos grupos, las medidas pueden tomarse sólo después, o bien antes y después de la aplicación del tratamiento. Así, se tienen los DGCNE con sólo medidas después y los DGCNE con medidas antes y después (figura 2.1). En cualquier caso, los sujetos pueden ser seleccionados de poblaciones diferentes, o bien se asignan grupos de sujetos a las condiciones control o de tratamiento (diseños de grupos intactos).

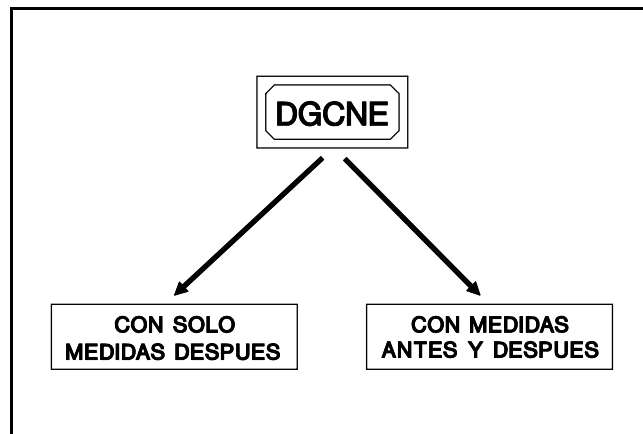


Figura 2.1. Clasificación de los diseños de grupo control no equivalente.

*Diseño de grupo control no equivalente con sólo medidas después.* Estructura cuasi-experimental de grupo control y de tratamiento. Los sujetos no son asignados aleatoriamente a los distintos grupos, con lo que una gran cantidad de variables pueden confundir la acción del tratamiento. En efecto, el principal inconveniente de los DGCNE con sólo medidas después radica en la dificultad de derivar inferencias causales sobre el impacto de la intervención. Por dicha razón, no es recomendable su uso.

*Diseño de grupo control no equivalente con medidas antes y después (DGCNE).* A diferencia del diseño anterior, al tomarse medidas antes de la aplicación del tratamiento, es posible verificar la equivalencia inicial de los grupos. Existen dos formatos básicos: a) diseños donde los sujetos son seleccionados de poblaciones que difieren en la medida de respuesta, y b) diseños de grupos intactos.

*Diseño de grupos no equivalentes con medidas antes y después.* Estructura de investigación cuasi-experimental con tres o más grupos de tratamiento, de acuerdo con los niveles de la variable independiente. Se trata, por tanto, de una extensión de los DGCNE.

Tanto los DGCNE como los diseños de grupos no equivalentes con medidas antes y después pueden ampliarse a situaciones donde se toman

varias medidas pre-tratamiento. Son los llamados *diseños de grupos no equivalentes con múltiples covariables*. Con este último diseño se incrementa la potencia del ACOVAR. En efecto, al incluir dos o más covariables, se ajustan los valores de la variable dependiente a diferentes fuentes de no equivalencia inicial consiguiendo, de esta forma, una estimación menos sesgada del efecto de los tratamientos.

### **2.1.1. Características del DGCNE con medidas antes y después**

- 1) Ausencia de aleatorización en la asignación de las unidades y variable de selección desconocida. Cabe la posibilidad que los grupos estén formados de manera natural antes de empezar la investigación (colegio, aula, distrito, hospital, etc.), o que los grupos se formen con sujetos seleccionados de dos poblaciones diferentes. En ambos casos, un grupo es asignado a la condición de tratamiento y otro grupo a la condición de control (grupo de comparación). Debido a la falta de aleatoriedad, es conveniente que los grupos sean similares en algunos aspectos relevantes, tales como características de los individuos y circunstancias contextuales.
- 2) De lo anterior se desprende la no equivalencia inicial de los grupos. Éstos presentan diferencias o sesgos antes de la aplicación del tratamiento capaces de contaminar su efecto.
- 3) Vulnerable a las amenazas contra la validez interna, especialmente aquellas que tienen su origen en la selección de las unidades. Sin embargo, dentro de la investigación aplicada, es el que posee mayor validez interna.
- 4) Las medidas pre-tratamiento permiten tener en cuenta o corregir las diferencias iniciales de los grupos mediante técnicas estadísticas (por ejemplo, el ACOVAR y el AVAR con puntuaciones de diferencia).

### 2.1.2. Técnicas de análisis

En un DGCNE, debido a la asignación no aleatoria, el efecto del tratamiento está sesgado, es decir, puede ser que los valores de la variable pre-tratamiento sean la causa de las diferencias en las medidas post-tratamiento. Por este motivo, el objetivo del investigador es estimar el efecto del tratamiento, controlando al máximo el efecto de las diferencias de selección. Existen seis técnicas de análisis fundamentales que pueden ser aplicadas a los datos de DGCNE de medidas antes y después (Arnau, 1994a, 1997; Judd y Kenny, 1981; Kenny, 1975a, 1979; Reichardt, 1979):

- A) AVAR de las puntuaciones después.
- B) ACOVAR.
- C) ACOVAR con corrección por la fiabilidad.
- D) AVAR con técnica de bloqueo.
- E) AVAR con puntuaciones de diferencia o ganancia.
- F) AVAR con puntuaciones de diferencia estandarizadas.

Respecto a cada uno de estos métodos de análisis, han habido diversas críticas. El argumento que subyace en la mayor parte de ellas es que "los supuestos de las técnicas analíticas no se satisfacen en las aplicaciones del DGCNE" (Mohr, 1982; p. 59).

#### **A) AVAR de las puntuaciones después**

Uno de los procedimientos de análisis más simples que se aplica a los DGCNE de medidas antes y después es el AVAR de las puntuaciones post-tratamiento. El objetivo es probar el impacto de la intervención, en el supuesto de la igualdad inicial de los grupos.

La técnica del AVAR consiste en comparar las variancias de los distintos grupos, a fin de comprobar si existen o no diferencias significativas entre las medias de los grupos. El modelo estructural viene representado por la siguiente expresión matemática:



$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij} \quad (1)$$

donde

$Y_{ij}$  =puntuación del  $i$ ésimo sujeto bajo el tratamiento  $j$ .

$\mu$  =media total.

$\alpha_j$  =efecto del  $j$ ésimo tratamiento.

$\varepsilon_{ij}$  =error asociado al sujeto  $i$  bajo el tratamiento  $j$ .

#### ESPECIFICACIONES DEL MODELO:

- 1) Existen dos posibles modelos, en función de que el experimentador elija los niveles a estudiar por cada factor (modelo de efectos fijos) o que los niveles sean muchos y se seleccionen al azar (modelo de efectos aleatorios). También pueden darse modelos mixtos que se caracterizan por contener componentes sistemáticos fijos y componentes aleatorios.
- 2) El componente  $\mu$  es constante en todas las observaciones del experimento.
- 3) El término  $\alpha_j$  es constante en todas las observaciones obtenidas dentro de cada grupo de tratamiento, bajo el supuesto del modelo de efectos fijos.
- 4) El componente  $\varepsilon_{ij}$  es independiente de  $\alpha_j$ .
- 5) Los errores del modelo y, por tanto, las observaciones son independientes. Además, los residuales tienen media cero y variancia constante  $\varepsilon_{ij} \sim NID(0, \sigma_\varepsilon^2)$ .
- 6) La estimación del efecto del tratamiento es:

$$\hat{\alpha}_E - \hat{\alpha}_C = \overline{Y_E} - \overline{Y_C}$$

donde el subíndice E se refiere al grupo experimental o de tratamiento y el subíndice C al grupo de control.

- 7) La variable dependiente es cuantitativa continua.
- 8) Los datos pueden contener información valiosa no recogida en el modelo matemático. En consecuencia, es importante tener en cuenta la posibilidad de omitir variables relevantes o incluir variables irrelevantes.
- 9) Existencia de tantos grupos como niveles de la variable independiente. Estos grupos deben constituir muestras aleatorias independientes en relación a la misma o distintas poblaciones.
- 10) Una correcta estimación de los parámetros requiere un número mayor de sujetos que valores de la variable independiente.
- 11) La variable dependiente debe presentar una distribución normal (o una aproximación a dicha distribución). De este modo, posee igual variancia en las distintas poblaciones de las que se han obtenido los grupos (condición de homocedasticidad).

El principal inconveniente de la aplicación del AVAR a las puntuaciones después en DGCNE es que no se tienen en cuenta las diferencias existentes en el pre-tratamiento y, por lo tanto, es posible que los resultados sean significativos debido a la diferencia inicial y no a la acción del tratamiento. De este modo, el AVAR posee una menor potencia, al compararlo con análisis alternativos donde se tienen en cuenta las diferencias de los grupos en las medidas antes (ACOVAR, AVAR con técnica de bloqueo y AVAR con puntuaciones de ganancia). Una solución al respecto consiste en aplicar, previamente, el AVAR a las puntuaciones de la variable antes. En caso de que no existan diferencias significativas, se presupone que los grupos son inicialmente equivalentes (Reichardt, 1979). Sin embargo, aunque se acepte la hipótesis de nulidad en este primer análisis, puede ocurrir que las medidas post-tratamiento presenten diferencias, aun cuando el tratamiento no sea efectivo. Ello es debido a una mayor variabilidad intra-grupo a medida que transcurre el tiempo (como consecuencia de la selección diferencial o del efecto de maduración).

## B) ACOVAR

El ACOVAR constituye una integración del AVAR y el AR. Su utilización reduce la variancia intra-grupal y, por lo tanto, aumenta la precisión del contraste estadístico, en comparación con el análisis común y elemental de la variancia. De acuerdo con el ACOVAR, se ajustan los datos del post-tratamiento, mediante el modelo de la regresión, a los valores de una variable concomitante o covariable. De esta manera, se obtienen unos datos ajustados a los que se aplica el correspondiente AVAR. En definitiva, se trata de reducir, en la medida de lo posible, el sesgo causado por las diferencias entre los grupos antes de la aplicación del tratamiento.

En el modelo matemático del ACOVAR (ecuación 2) se incluye la variable pre-tratamiento en forma de regresión lineal. Por tanto, el único término nuevo que aparece, en relación con el AVAR, es el coeficiente  $\beta$  :

$$Y_{ij} = \mu + \alpha_j + \beta(X_{ij} - \overline{X_{..}}) + \varepsilon'_{ij} \quad (2)$$

donde

$Y_{ij}$  = puntuación del sujeto  $i$  bajo el tratamiento  $j$ .

$\mu$  = media de todas las observaciones.

$\alpha_j$  = efecto del  $j$ ésimo tratamiento.

$\beta$  = coeficiente de la regresión lineal intra-grupo de  $Y$  sobre  $X$ .

$X_{ij}$  = puntuación en la covariable del sujeto  $i$  en el grupo  $j$ .

$\overline{X_{..}}$  = media total de la covariable.

$\varepsilon'_{ij}$  = error asociado al sujeto  $i$  bajo el tratamiento  $j$ .

### ESPECIFICACIONES DEL MODELO:

Además de las condiciones que exige el AVAR, la correcta aplicación del ACOVAR añade otras especificaciones.

- 1) Los errores asociados con cada uno de los valores de la variable dependiente tienen distribución normal y son independientes entre

sí  $(\varepsilon'_{ij} \sim NID(0, \sigma_\varepsilon^2))$ .

2) La estimación del efecto del tratamiento es:

$$\hat{\alpha}_E - \hat{\alpha}_C = (\overline{Y}_E - \overline{Y}_C) - \hat{\beta}(\overline{X}_E - \overline{X}_C)$$

3) Al introducir el pre-tratamiento en el modelo como una desviación de la media total  $(\overline{X}_{ij} - \overline{X}_{..})$ , la estimación del efecto principal del tratamiento es:

$$\hat{\alpha}_E - \hat{\alpha}_C = (\overline{Y}_E - \overline{Y}_C) - \hat{\beta}_E(\overline{X}_E - \overline{X}_{..}) + \hat{\beta}_C(\overline{X}_C - \overline{X}_{..})$$

4) Tanto la variable dependiente como la covariable son cuantitativas.

5) Linealidad de la regresión de  $Y$  sobre  $X$ .

6) Homogeneidad de los coeficientes de la regresión.

7) Los tratamientos no afectan a la covariable. Este supuesto se cumple siempre que las medidas de la covariable se tomen antes de la aplicación del tratamiento.

El supuesto más importante consiste en verificar la homogeneidad de las pendientes de regresión de ambos grupos (control y tratamiento), lo que nos lleva a esperar que la proporción de cambio producida por la interacción entre la variable concomitante y la de tratamiento sea no significativa (Edwards, 1984; Rogosa, 1980a).

Un aspecto importante a considerar es que a mayor relación entre la variable dependiente y la covariable, mayor reducción de la variancia del error con la que se contrasta el efecto del tratamiento. Por otra parte, el ACOVAR pierde precisión a medida que se incrementan las diferencias iniciales entre los grupos.

La correcta aplicación del ACOVAR requiere que la covariable y la

variable dependiente midan la misma dimensión o constructo. Por otra parte, cuando la covariable contiene errores de medida (aspecto muy frecuente en la mayoría de situaciones de investigación en ciencias del comportamiento), se puede producir un sesgo en la estimación del efecto del tratamiento (Arnau, 1994a, 1997; Ato, 1995a; Huitema, 1980; Mohr, 1982; Pedhazur y Schmelkin, 1991). Otra dificultad, que atenta contra la validez interna, es la regresión hacia la media, muy común cuando la variable de pre-tratamiento es medida con error.

Ante estos inconvenientes, se han buscado métodos alternativos de análisis. Entre ellos destacan el ACOVAR con puntuaciones verdaderas o corregidas por la fiabilidad, el AVAR con técnica de bloqueo, el AVAR con puntuaciones de diferencia o ganancia y el AVAR con puntuaciones de diferencia estandarizadas.

### C) ACOVAR con corrección por la fiabilidad

Cuando la fiabilidad de las medidas de la covariable no es exacta, se produce un sesgo negativo en las estimaciones de los coeficientes de regresión (Lord, 1960; Porter y Chibucos, 1974; Reichardt, 1979). Es decir, los errores de medida en la variable pre-tratamiento llevan a una subestimación o atenuación del coeficiente de la regresión (la estimación de la pendiente es más pequeña de lo que le correspondería, si la medida hubiese sido fiable).

Con objeto de corregir la atenuación en la estimación del coeficiente de la regresión, se desarrolla el procedimiento del ACOVAR con puntuaciones verdaderas de la covariable (Porter, 1967). La estimación de dichas puntuaciones se obtiene a través de la siguiente expresión:

$$\hat{T}_{ij} = \overline{X}_{.j} + r_{xx}(X_{ij} - \overline{X}_{.j}) \quad (3)$$

donde

$\hat{T}_{ij}$  = puntuación verdadera estimada del *i*ésimo sujeto bajo el grupo *j*.

$X_{ij}$  = puntuación en la covariable del sujeto *i* bajo el grupo *j*.

$\overline{X}_{.j}$  = media de valores de la covariable del grupo *j*.

$r_{xx}$  = fiabilidad estimada de la covariable.

La ecuación (3) incluye una estimación del coeficiente de fiabilidad de la variable antes, con lo que se plantea el problema del cálculo de dicha fiabilidad. Cuando se dispone de una sola covariable se sugiere utilizar la correlación intra-grupo pre-post tratamiento conjunta (ecuación 4), y en caso de existir dos o más covariables, algunos de los procedimientos para calcular la fiabilidad son la correlación entre formas paralelas y la correlación test-retest (Huitema, 1980).

$$r_{xy(ij)} = \frac{SP_{xy(ij)}}{\sqrt{(SC_{x(ij)})(SC_{y(ij)})}} \quad (4)$$

donde

$SP_{xy(ij)}$  = Suma de productos intra-grupo conjunta de las variables antes y después.

$SC_{x(ij)}$  = Suma de cuadrados intra-grupo conjunta de la variable antes.

$SC_{y(ij)}$  = Suma de cuadrados intra-grupo conjunta de la variable después.

Una vez calculada la estimación de la fiabilidad, se estiman los valores verdaderos de la covariable (ecuación 3) y, por último, se lleva a cabo el ACOVAR de la variable post-tratamiento con las puntuaciones corregidas por la fiabilidad. El modelo estructural se muestra en la expresión (5) y sus especificaciones son las propias del ACOVAR.

$$Y_{ij} = \mu + \alpha_j + \beta(T_{ij} - \bar{T}_{..}) + \varepsilon'_{ij} \quad (5)$$

#### D) AVAR con técnica de bloqueo

La técnica de bloqueo permite agrupar, a partir de la matriz de datos inicial del diseño, los sujetos de acuerdo con unos intervalos de la variable pre-tratamiento. En otras palabras, los bloques se forman a partir de individuos que tienen puntuaciones similares en la covariable. Un caso extremo de la técnica de bloqueo es la de apareo, donde cada par de sujetos constituye un bloque.

La variable de bloques se incorpora en el modelo estructural junto con la variable de tratamiento y la interacción entre ambas (bloques x tratamiento).

Así, el modelo estructural del AVAR con técnica de bloqueo viene expresado por la siguiente fórmula:

$$Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk} \quad (6)$$

donde

$Y_{ijk}$  = puntuación del  $i$ ésimo sujeto bajo el  $j$ ésimo tratamiento y el  $k$ ésimo bloque.

$\mu$  = media de todas las observaciones.

$\alpha_j$  = efecto del tratamiento  $j$ .

$\beta_k$  = efecto del bloque  $k$ .

$(\alpha\beta)_{jk}$  = efecto de la interacción entre la variable de bloques y la de tratamiento.

$\varepsilon_{ijk}$  = error asociado al sujeto  $i$  bajo el tratamiento  $j$  y el bloque  $k$ .

#### ESPECIFICACIONES DEL MODELO:

En la técnica del bloqueo se mantienen los supuestos propios del AVAR. Sin embargo, consideramos oportuno matizar algunos aspectos.

- 1)  $\varepsilon_{ijk} \sim NID(0, \sigma_\varepsilon^2)$
- 2) La covariable es categórica.
- 3) La variable de bloques debe estar correlacionada con la variable dependiente.
- 4) Cuando se dispone de un solo sujeto por casilla, la interacción entre la variable de bloques y la de tratamiento se confunde con el error y, en consecuencia, dicha interacción se utiliza como término de contrastación. En cambio, cuando cada celda está formada por dos o más sujetos, se puede verificar la significación de la interacción.

El efecto de tratamiento se estima a partir de la diferencia entre los grupos dentro de cada bloque (Reichardt, 1979). Conforme aumenta el número de bloques, la precisión del procedimiento de bloqueo se aproxima a la del ACOVAR (Cochran, 1968).

Respecto a la existencia de desproporcionalidad en las celdas o casillas, un método adecuado es el de las medias no ponderadas que trata las medias de las casillas como si estuvieran basadas en  $n$ 's iguales (Arnau, 1994a, 1997). Sin embargo, este procedimiento se aconseja cuando las  $n$ 's no son muy desproporcionadas.

### E) AVAR con puntuaciones de diferencia

El análisis basado en puntuaciones de cambio o ganancia consiste en hallar, en cada uno de los casos, la diferencia entre las puntuaciones después y los valores antes, y a continuación se prueba la significación estadística a través del AVAR.

El modelo estructural viene representado por la fórmula (7). Las puntuaciones de diferencia son consideradas como la variable dependiente.

$$Y_{ij} - X_{ij} = \mu + \alpha_j + \varepsilon_{ij} \quad (7)$$

donde

$Y_{ij}$  = puntuación del  $i$ ésimo sujeto bajo el tratamiento  $j$ .

$X_{ij}$  = puntuación en la covariable del sujeto  $i$  bajo el grupo  $j$ .

$\mu$  = media de todas las observaciones.

$\alpha_j$  = efecto del  $j$ ésimo tratamiento.

$\varepsilon_{ij}$  = error asociado al sujeto  $i$  bajo el tratamiento  $j$ .

### ESPECIFICACIONES DEL MODELO:

1)  $\varepsilon_{ij} \sim NID(0, \sigma_\varepsilon^2)$

2) La estimación del efecto del tratamiento es:

$$\hat{\sigma}_E - \hat{\sigma}_C = (\overline{Y_E} - \overline{Y_C}) - (\overline{X_E} - \overline{X_C})$$



- 3) Las tasas de crecimiento de la variable de asignación deben ser semejantes.

Un correcto análisis de las puntuaciones de cambio requiere que se cumpla el tercer supuesto, denominado de estacionariedad (Judd y Kenny, 1981). Esta condición revela la ausencia de interacción entre selección (variable de asignación) y maduración o tasa de crecimiento (Ato, 1995a; Campbell, 1988; Campbell y Stanley, 1963; Kenny, 1975a). Si no se cumple esta condición, al aplicar el análisis de las puntuaciones de ganancia, se produce un efecto que se interpreta equivocadamente como debido al tratamiento, pero en realidad es debido a la tasa de crecimiento diferencial.

Una interacción selección-maduración hace que los sujetos con puntuaciones altas en la covariable obtengan mejoras de mayor cuantía en el post-tratamiento. En otras palabras, los sujetos con niveles altos poseen una tasa de maduración más alta que los sujetos con niveles bajos (Arnau, 1994a), de modo que las medias de datos de diferencia de grupos con medias pre-tratamiento distintas no son iguales. Este fenómeno, producido como consecuencia del paso del tiempo, se denomina hipótesis de dispersión en abanico (*fan-spread hypothesis*) debido a la peculiar forma que toma (Bryk y Weisberg, 1977, pp. 956-957; Huitema, 1980, p. 328; Judd y Kenny, 1981, p. 121). Dicha dispersión origina un incremento de la variancia intra y entre-grupo.

Un aspecto a destacar es que el AVAR con puntuaciones de ganancia es más preciso que el ACOVAR y el análisis de bloques, siempre y cuando  $\beta = 1$  (Cox, 1957; Feldt, 1958; Reichardt, 1979). Por otra parte, el análisis de las puntuaciones de diferencia tiene mayor potencia que el AVAR cuando  $0 < 1/\beta < 2$  (Cox, 1957).

## **F) AVAR con puntuaciones de diferencia estandarizadas**

El método de los valores de diferencia no es adecuado cuando se da el

fenómeno de la dispersión en abanico. En otras palabras, cuando la variancia entre e intra-grupo aumenta como consecuencia del transcurso del tiempo y de los procesos de maduración (Cook y Campbell, 1979; Huitema, 1980; Judd y Kenny, 1981). En este caso, la diferencia entre las variancias de las variables antes y después produce un sesgo en la estimación del efecto del tratamiento. Este problema puede ser solventado mediante algún procedimiento que permita homogeneizar las variancias, como por ejemplo la transformación de los datos a través de la estandarización.

Los valores de diferencia estandarizados se obtienen al dividir, en primer lugar, las puntuaciones pre-tratamiento y post-tratamiento por la correspondiente desviación estándar intra-grupo combinada (Judd y Kenny, 1981) y, posteriormente, se calculan las diferencias de los datos estandarizados.

El modelo estructural del AVAR con puntuaciones de diferencia estandarizadas viene expresado de la siguiente forma:

$$\frac{Y_{ij}}{\sigma_{Y(ij)}} - \frac{X_{ij}}{\sigma_{X(ij)}} = \mu + \alpha_j + \varepsilon_{ij} \quad (8)$$

donde

$Y_{ij}$  = puntuación del  $i$ ésimo sujeto bajo el tratamiento  $j$ .

$X_{ij}$  = puntuación en la covariable del sujeto  $i$  bajo el grupo  $j$ .

$\sigma_{Y(ij)}$  = desviación estándar intra-grupo combinada de la variable después.

$\sigma_{X(ij)}$  = desviación estándar intra-grupo combinada de la variable antes.

$\mu$  = media de todas las observaciones.

$\alpha_j$  = efecto del  $j$ ésimo tratamiento.

$\varepsilon_{ij}$  = error asociado al sujeto  $i$  bajo el tratamiento  $j$ .

La desviación estándar intra-grupo combinada, tanto de la variable antes como después es:

$$\sigma_{(ij)} = \sqrt{\frac{SC_C + SC_E}{n_C + n_E - 2}} \quad (9)$$

donde  $n_C$  y  $n_E$  simbolizan la cantidad de datos del grupo control y tratamiento,

respectivamente.

#### ESPECIFICACIONES DEL MODELO:

- 1) Las puntuaciones de diferencia estandarizadas tienen una distribución normal.
- 2) La estimación del efecto del tratamiento es:

$$\hat{\sigma}_E - \hat{\sigma}_C = (\bar{Y}_E - \bar{Y}_C) - \frac{\sigma_{Y(ig)}}{\sigma_{X(ig)}} (\bar{X}_E - \bar{X}_C)$$

En caso de que los valores de ganancia no se ajusten a un patrón de dispersión en abanico, el análisis de las puntuaciones de diferencia estandarizadas estará sesgado.

### **2.2. Diseños de discontinuidad en la regresión (DDR)**

El DDR es una estructura cuasi-experimental cuyo formato básico consta de una medida o variable antes y una medida después que refleja el efecto de tratamiento. En esta clase de diseño, el criterio de asignación de las unidades a los grupos es conocido. Una descripción más precisa del DDR se da a través de su representación gráfica (figura 2.2).

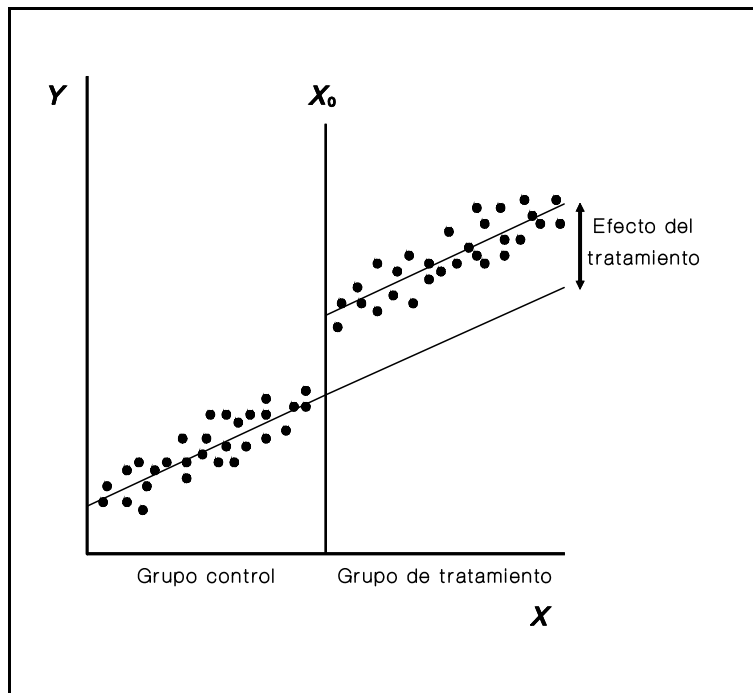


Figura 2.2. Representación gráfica del DDR.

En la figura 2.2 el eje de ordenadas se refiere al rango de valores de la variable post-tratamiento ( $Y$ ) y el eje de abscisas representa los valores de la variable pre-tratamiento ( $X$ ), con base a la cual se define un punto de corte ( $X_0$ ) que determina la pertenencia de los sujetos al grupo control (los que muestran valores inferiores a dicho punto) o al grupo de tratamiento (los que se sitúan por encima del punto de corte). Las líneas continuas representan las rectas de la regresión de  $Y$  sobre  $X$  para los grupos control y tratamiento, y la línea punteada muestra como sería la recta de regresión del grupo tratado si la intervención no hubiese tenido ningún efecto. La magnitud del desplazamiento de la línea de regresión en el grupo de tratamiento indica la amplitud del efecto de intervención.

### 2.2.1. Características del DDR

- 1) El criterio de formación de los grupos es explícito y conocido, aunque no aleatorio. Los sujetos se asignan a los grupos de

tratamiento y control en función de un punto de corte en la variable pre-tratamiento, de manera que aquellos sujetos con valores por debajo o encima de  $X_0$  son asignados a uno u otro grupo. Por motivos éticos, el tratamiento se aplica a aquellos sujetos que más lo necesitan.

- 2) La variable de asignación puede ser una medida pre-tratamiento de la misma variable después, o bien de cualquier otra variable relevante para el objeto de estudio registrada antes de administrar el tratamiento.
- 3) En caso de que el efecto del tratamiento sea nulo, y asumiendo que la regresión de la variable  $Y$  sobre la  $X$  es lineal, cabe esperar que la línea de regresión iniciada en el grupo control se extienda al grupo de tratamiento.
- 4) La discontinuidad de las rectas en el punto de corte refleja el efecto del tratamiento.
- 5) Se pueden analizar tanto efectos principales (cuando se produce un desplazamiento vertical de las líneas de la regresión en el punto de corte) como efectos de interacción (cuando el tratamiento está asociado a la variable antes). Trochim (1984) distingue cinco posibles efectos (figura 2.3):

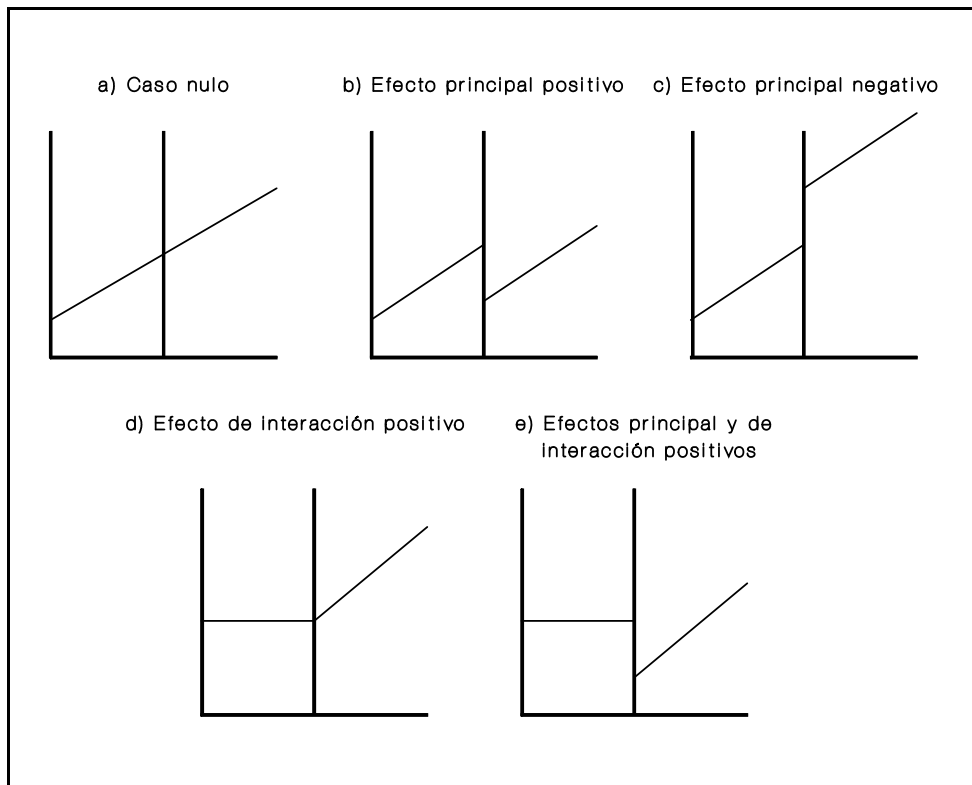


Figura 2.3. Patrones hipotéticos de las líneas de regresión (Trochim, 1984; p. 54).

- 6) La estimación de los efectos del tratamiento no está sesgada, incluso cuando la covariable contiene errores de medida (Cappelleri *et al.*, 1991; Pedhazur y Schmelkin, 1991; Reichardt, 1979; Reichardt *et al.*, 1995) y también cuando existe efecto de interacción (Reichardt *et al.*, 1995; Rubin, 1977; Trochim *et al.*, 1991). Ello se debe a que el criterio de asignación es conocido y, por lo tanto, en el ACOVAR el efecto del tratamiento se puede aislar del error de medida de la covariable. En este sentido, el DDR funciona igual que el diseño experimental clásico. Sin embargo, las publicaciones acerca de este tema han sido, hasta el momento, confusas y contradictorias. Así, por ejemplo, Stanley (1991) y Stanley y Robinson (1990) muestran que en situaciones donde hay errores de medida o interacción tratamiento-covariable, se obtienen estimadores sesgados del efecto de tratamiento.
- 7) Las aplicaciones del DDR son limitadas debido a que pocas veces las

reglas de selección están suficientemente definidas. Sólo dentro del ámbito educativo se constata un uso regular del DDR (Trochim, 1984).

### 2.2.2. Técnicas de análisis

En el análisis de datos del DDR se utiliza el ACOVAR, bajo el supuesto de una relación lineal y aditiva entre la variable antes y la de tratamiento. Su modelo estructural es el mostrado en el DGCNE (ecuación 2). Otra técnica de análisis, también útil en los DDR, es el ARM, cuya versión más simplificada se muestra a continuación. En su exposición, no vamos a insistir, como es obvio, que se asume la linealidad del modelo de la regresión y aditividad de sus componentes.

#### A) ARM con rectas de regresión paralelas

El modelo lineal de la regresión se representa a través de la siguiente ecuación general:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 T_i + \varepsilon_i \quad (10)$$

donde

$Y_i$  = valor del sujeto  $i$  en la variable post-tratamiento.

$X_i$  = medida del individuo  $i$  en la variable pre-tratamiento.

$T_i$  = variable *dummy* (grupo control=0 y grupo de tratamiento=1).

$\beta_0$  = intercepción de la recta de ajuste.

$\beta_1$  = pendiente de la línea de regresión.

$\beta_2$  = diferencia entre grupo control y de tratamiento.

$\varepsilon_i$  = variación residual del individuo  $i$

ESPECIFICACIONES DEL MODELO:

1)  $\varepsilon_i \sim NID(0, \sigma_\varepsilon^2)$

2) La variable pre-tratamiento y la variable dependiente son

cuantitativas.

- 3) Tanto el grupo control como el de tratamiento deben tener una misma distribución en la variable antes.
- 4) Homogeneidad de las pendientes de regresión.

El análisis estadístico ajusta, en primer lugar, la recta de regresión de cada grupo (control y tratamiento). La diferencia de las intercepciones de ambas líneas de regresión en el punto de corte es la medida del efecto de tratamiento. Así, asumiendo el modelo de la ecuación (10) y que el tratamiento es efectivo, cabe esperar que las dos rectas de regresión tengan igual pendiente ( $\beta_1$ ), y que el parámetro  $\beta_2$  asociado a la variable *dummy* sea distinto de cero. En caso contrario ( $\beta_2=0$ ), una sola línea de regresión une el grupo control y el de tratamiento y, en consecuencia, se concluye que la intervención no es eficaz.

### **B) ARM con pendientes no homogéneas**

Para que el modelo de la ecuación (10) sea válido, debe asumirse que las pendientes de las rectas de regresión son paralelas. Por el contrario, se tendría que introducir un factor que recoja la interacción  $XT$  (ecuación 11). Esta nueva variable permite verificar el supuesto de homogeneidad de las pendientes.

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 T_i + \beta_3 X_i T_i + \varepsilon_i \quad (11)$$

donde

$Y_i$  = medida del sujeto  $i$  en la variable post-tratamiento.

$X_i$  = valor del individuo  $i$  en la variable antes.

$T_i$  = variable *dummy* que recoge la pertenencia del dato al grupo control o de tratamiento.

$\beta_0$  = intercepción de la línea de regresión.

$\beta_1$  = pendiente de la recta de ajuste.

$\beta_2$  = efecto de tratamiento.

$\beta_3$  = diferencia entre las pendientes del grupo control y de tratamiento.

$\varepsilon_i$  = residual.

Cuando existe interacción, el efecto del tratamiento varía para los



distintos valores de la variable antes. Un ejemplo de interacción  $XT$  sería el caso de una pendiente más pronunciada en el grupo de tratamiento. Ello significa que los sujetos con puntuaciones altas en la covariable se benefician en mayor medida del tratamiento que los que puntúan bajo.

El objetivo básico del DDR es la comparación de las dos ecuaciones de regresión en el punto de corte para estimar el efecto del tratamiento. Sin embargo, cuando la interacción no es nula se presenta un problema: el desplazamiento de las rectas de regresión no es constante y, por lo tanto, es difícil decidir en qué punto debe medirse dicho desplazamiento. Una solución, dada por Trochim (1984), consiste en llevar a cabo el análisis con los datos de diferencia de los valores pre-tratamiento respecto al punto de corte ( $X' = X - X_0$ ). Esta transformación tiene el efecto de mover el valor de corte al punto de la intercepción  $Y$ , lo cual mejora la interpretabilidad de los coeficientes.

### 3. DISEÑOS LONGITUDINALES

#### 3.1. Diseños de series temporales interrumpidas (DSTI)

Un procedimiento que permite mejorar los diseños cuasi-experimentales consiste en tomar múltiples registros, del mismo sujeto, a lo largo del tiempo. Por este motivo, se denominan *diseños de series temporales*. Existen dos modalidades básicas de este tipo de diseño (Arnau, 1995b): a) el DSTI simple, y b) el DSTI de grupo control no equivalente.

*DSTI simple*. Conjunto de datos registrados secuencialmente en una serie discreta de puntos u observaciones de una sola unidad observacional, antes y después de la presencia de un evento o aplicación de un tratamiento. El objetivo del DSTI es verificar si la introducción de la intervención produce cambios de nivel y/o pendiente en la serie.

*DSTI de grupo control no equivalente*. Incorporación al DSTI simple de un grupo control lo más parecido posible al grupo tratado, con observaciones

registradas al mismo tiempo y bajo las mismas circunstancias que el grupo de tratamiento.

### **3.1.1. Características del DSTI**

- 1) Modelo prototípico de investigación longitudinal.
- 2) Los puntos de tiempo están serialmente secuenciados, antes y después de la intervención.
- 3) La unidad observacional puede ser un individuo o un colectivo.
- 4) Existe un sólo dato en cada uno de los puntos de tiempo y una clara línea divisoria en la aplicación del tratamiento.
- 5) Se asume que las observaciones están igualmente espaciadas a fin de interpretar correctamente los resultados.
- 6) El número de registros ha de ser suficiente (50 o más observaciones por fase) con objeto de poder aplicar el método de análisis de Box y Jenkins (1970).

### **3.1.2. Análisis de series temporales: modelos ARIMA**

En la exposición del análisis de series temporales que se presenta a continuación nos basaremos en los trabajos de Arnau (1995d, 1996), Box y Jenkins (1970), Glass *et al.* (1975), Gottman (1981) y McCleary y Hay (1980) y Vallejo (1996).

Desde el punto de vista estadístico, las *series temporales empíricas* se definen como realizaciones de un proceso estocástico subyacente o modelo teórico. Así, el objetivo del analista de los datos es inferir, a partir de una realización, el mecanismo generador de la serie o proceso estocástico latente, mediante un modelo autorregresivo integrado de medias móviles (ARIMA). En definitiva, se trata de identificar el orden de los componentes sistemáticos

de una serie temporal.

Debido a la propia estructura de los datos (la misma unidad es observada a lo largo de una secuencia de puntos de tiempo), las observaciones y los errores de una serie temporal están autocorrelacionados. Con objeto de determinar esta estructura de dependencia existen dos instrumentos diagnósticos: la *función de autocorrelación* (FAC) y la *función de autocorrelación parcial* (FACP). El valor que puede tomar una autocorrelación indica el grado en que una determinada observación del proceso se halla correlacionada con otra en un retardo  $k$ . De este modo, la  $FAC(0)$  es la correlación de la serie original consigo misma que, obviamente, sería igual a la unidad; la  $FAC(1)$  es la correlación de la serie original con la retardada (la serie se desplaza un lugar hacia adelante), y así sucesivamente se calculan el resto de valores de la función de autocorrelación.

Las representaciones gráficas de la FAC y la FACP, denominadas autocorrelogramas, aportan la información necesaria para conocer las propiedades del mecanismo generador de una serie temporal. Tanto la FAC como la FACP nos permiten detectar los componentes sistemáticos de una serie temporal. Así, cualquier valor significativo de estas dos funciones es un indicio de dependencia. En caso que no haya ninguna autocorrelación significativa, se concluye que la serie ha sido generada por un proceso formado por una secuencia de variables aleatorias  $\{a_t\}$ , mutuamente independientes e idénticamente distribuidas, con media cero y variancia constante,  $NID(0, \sigma_a^2)$ . Este proceso se denomina de *ruido blanco*.

Los modelos ARIMA están formados por tres componentes que pueden ser regulares o cíclicos: autorregresivo (AR), de integración (I) y de medias móviles (MA). Los parámetros asociados al modelo indican el orden en que actúa cada componente. De este modo,  $p$  es el orden del componente AR,  $d$  el nivel del componente I, y  $q$  el orden del componente MA. Por lo general, los procesos autorregresivos y de medias móviles más comunes en ciencias del comportamiento son los de primer y segundo orden.

*Componente de integración, ARIMA (0,d,0)*. La serie temporal refleja el efecto acumulativo de algún proceso o la presencia de tendencia, que puede

ser estocástica o determinista. Estos procesos no estacionarios, donde el nivel de la media cambia con el tiempo, se denominan de recorrido aleatorio (*random walk*).

Un proceso de recorrido aleatorio con componente de tendencia estocástico es aquel cuyo valor de la variable  $Y$  en el punto  $t$  es igual a su valor en el punto  $t-1$  más un impulso aleatorio  $\{a_t\}$ :

$$Y_t = Y_{t-1} + a_t \quad (12)$$

La expresión matemática anterior es equivalente a:

$$Y_t - Y_{t-1} = a_t \quad (13)$$

donde el proceso de ruido blanco tiene la siguiente propiedad:  $a_t \sim NID(0, \sigma_a^2)$ .

Cuando el nivel medio de la serie es función fija del tiempo, entonces el proceso de recorrido aleatorio posee, además, un componente de tendencia determinista:

$$Y_t = Y_{t-1} + \theta_0 + a_t \quad (14)$$

donde  $\theta_0$  es la pendiente y  $a_t \sim NID(0, \sigma_a^2)$ .

En cualquiera de los procesos de recorrido aleatorio, el componente I se modeliza mediante diferenciación de la serie, y el residual obtenido puede ser un proceso de ruido blanco o bien presentar otra clase de dependencia serial que debe ser modelada (proceso autorregresivo o de media móvil).

*Proceso autorregresivo, ARIMA (p,0,0)*. La observación actual se define por la suma ponderada de una cantidad finita de  $p$  observaciones precedentes más un impulso aleatorio independiente  $\{a_t\}$ :

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + a_t \quad (15)$$

donde  $y_t$  es definida en términos de desviación de la media ( $y_t = Y_t - \mu_t$ );  $a_t \sim NID(0, \sigma_a^2)$ , y los parámetros  $\phi_p$  son estadísticamente significativos y cumplen la condición de estacionariedad ( $-1 < \phi_p < +1$ ).

La característica básica de un proceso AR es que el pasado incide en el presente con un determinado peso, que decrece conforme nos alejamos históricamente de la observación actual.

*Proceso de media móvil, ARIMA (0,0,q)*. La observación  $y_t$  es descrita por una suma del impulso actual y los impulsos pasados ponderados, siendo estos últimos retardados en  $q$  períodos:

$$y_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (16)$$

donde  $a_t \sim NID(0, \sigma_a^2)$  y, de forma análoga a los parámetros autorregresivos, los parámetros de medias móviles deben ser estadísticamente significativos y cumplir la condición de invertibilidad ( $-1 < \theta_q < +1$ ).

Además de los modelos anteriores, pueden darse, dentro de un contexto psicológico y social, procesos donde estén presentes, de forma simultánea, componentes AR y MA. En cualquiera de los casos, los parámetros deben hallarse fuera del círculo de la unidad.

El procedimiento general a seguir en la construcción de un modelo autorregresivo integrado de medias móviles se presenta en la figura 3.1. Los tres pasos fundamentales son el de identificación del modelo ARIMA, estimación de los parámetros, y verificación de la adecuación del modelo o diagnóstico (McCleary y Hay, 1980).

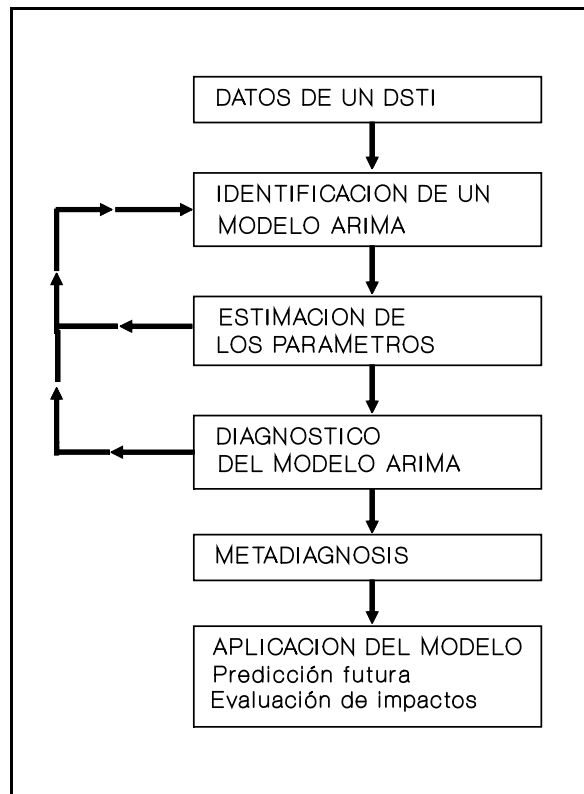


Figura 3.1. Proceso de construcción de un modelo ARIMA (modificada de Arnau, 1995d; p. 109).

*Identificación del modelo ARIMA.* A partir de los autocorrelogramas de la FAC y FACP se observa si existe algún proceso no estacionario. En caso afirmativo, deberá diferenciarse la serie. Una vez modelado el proceso de tendencia, se analizan los autocorrelogramas de los residuales a fin de detectar la presencia de dependencia serial (algún proceso autorregresivo o de media móvil). Cualquier decisión al respecto se toma comparando los patrones de las FAC y FACP.

*Estimación de los parámetros del modelo.* Los parámetros estimados deben ser estadísticamente significativos y han de situarse dentro de los límites de estacionariedad/invertibilidad. Cuando no se cumpla alguno de estos dos criterios, se regresa de nuevo a la fase de identificación y se propone otro modelo tentativo.

*Diagnóstico del modelo ARIMA.* Los residuales del modelo han de seguir un proceso de ruido blanco. Por el contrario, debe regresarse a la fase de identificación.

*Metadiagnosis.* En caso de que se hayan encontrado dos modelos plausibles, existen una serie de estadísticos, como por ejemplo el cuadrado medio residual, que miden la bondad de ajuste de los modelos propuestos. El modelo que posea el valor más pequeño es el que explica mejor la serie, ya que minimiza la variancia de los residuales.

*Aplicación de los modelos ARIMA.* La modelación de las series temporales puede usarse para predecir los valores no observados, y también permite evaluar el impacto de los tratamientos. En este segundo caso, se modela previamente la serie en términos de sus componentes de dependencia serial y, a continuación, se añade a la serie un modelo de transferencia o de intervención que explique el posible efecto de una variable de tratamiento (efecto abrupto o gradual). El modelo global de la evaluación del impacto viene expresado por la siguiente ecuación:

$$Y_t = f(I_t) + N_t \quad (17)$$

donde  $f(I_t)$  es el componente de intervención y  $N_t$  el componente ARIMA.

### **3.2. Diseños de medidas repetidas (DMR)**

El interés por los DMR, en contextos aplicados, radica en la posibilidad de tomar medidas de una variable en dos o más puntos del tiempo. Así, la estrategia de medidas repetidas permite estudiar las tendencias que presentan los datos en función del tiempo (Arnau, 1995e; Bock, 1975, 1979; Stevens, 1986). Arnau (1997) propone una clasificación general del DMR en función de la cantidad de grupos (figura 3.2).

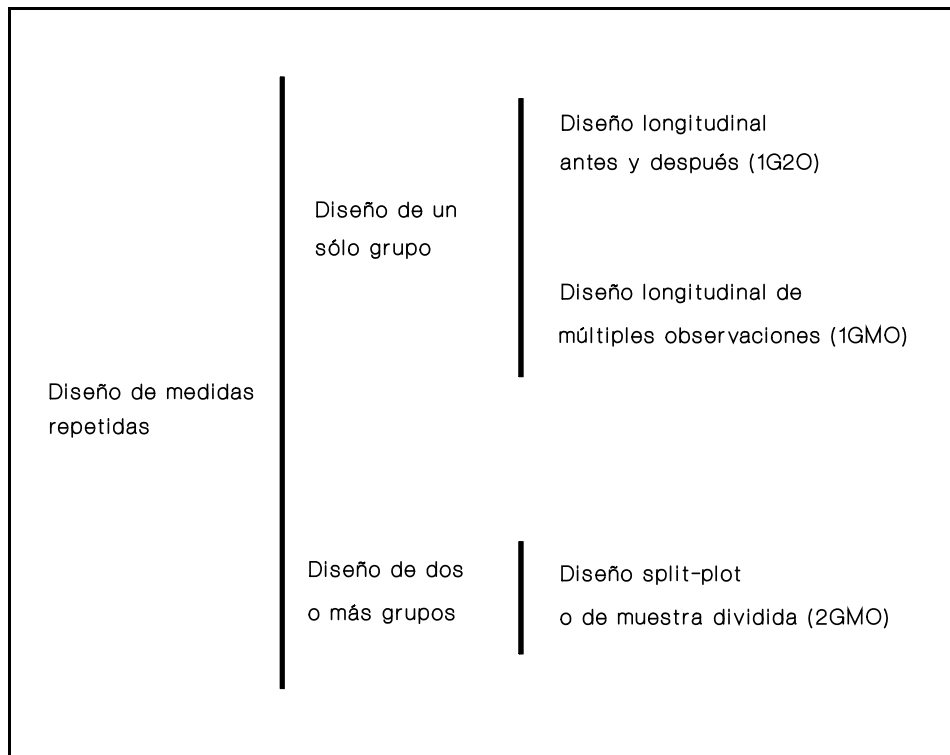


Figura 3.2. Clasificación del DMR (Arnau, 1997).

*Diseño longitudinal antes y después (1G2O).* Esquema prototípico de investigación longitudinal, cuyo propósito es el estudio del cambio de un mismo grupo de sujetos entre dos ocasiones de observación (como consecuencia de algún hecho circunstancial, tratamiento, o por el simple paso del tiempo).

*Diseño longitudinal de múltiples observaciones (1GMO).* Un solo grupo de sujetos es medido, respecto a una misma variable dependiente, en diferentes intervalos de tiempo, con objeto de determinar la curva de crecimiento.

*Diseño split-plot o de muestra dividida (2GMO).* Los sujetos se agrupan de acuerdo con los niveles de una variable de clasificación o categórica (variable pronóstica), tomándose medidas repetidas de cada uno de los sujetos en las distintas ocasiones de observación. La variedad más simple consta de dos grupos y una sola variable de registro.



### 3.2.1. Características del DMR

- 1) El DMR incorpora varios sujetos o unidades individuales que pueden formar parte de un solo grupo o de distintos grupos, en función de si se utiliza una variable de clasificación.
- 2) Cada unidad individual se registra a lo largo de una serie reducida de ocasiones.
- 3) Los registros pueden ser una respuesta a un tratamiento previo o una medida conductual.
- 4) Instrumento apropiado para el estudio del crecimiento o tendencias que presentan los datos en función del tiempo (modelación de las curvas de crecimiento) y evaluación de los procesos de cambio.

### 3.2.2. Técnicas de análisis del diseño longitudinal antes y después

El diseño 1G2O se ha utilizado básicamente como un procedimiento para la medida y evaluación del cambio. Con este objetivo, Goldstein (1979) propone dos modelos de análisis: los *modelos condicionales* y los *modelos incondicionales*.

#### A) Modelo condicional

El modelo condicional, comúnmente denominado modelo de la regresión, aplicado a datos de diseños 1G2O viene expresado por la ecuación (18). Mediante los modelos condicionales se plantea cuáles son los sujetos que ganan más o menos, y cuál es la causa de dichos cambios.

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon \quad (18)$$

donde

$Y$  = valor observado en la segunda ocasión de observación.

$X_1$  = medida de la primera ocasión.

$\beta_0$  = intercepción de la línea.

$\beta_1$  = pendiente.

$\varepsilon$  = término de error que incluye un conjunto de variables distintas de  $X_1$  que actúan de forma aleatoria sobre  $Y$ .

Cuando aplicamos un modelo condicional, se asume que las medidas de la primera ocasión constituyen una variable fija ( $X_1$ ) y se opera con la distribución de medidas de la segunda ocasión ( $Y$ ).

## **B) Modelo incondicional**

A través de los modelos incondicionales o modelos referidos al tiempo, se especifica el cambio en términos de diferencias entre las medias de las dos ocasiones ( $\bar{d} = \bar{Y} - \bar{X}$ ). Respecto al cambio individual, que constituye el centro de interés de los datos longitudinales, se tiene que el valor de diferencia entre las puntuaciones es  $d = Y - X$ , en cada uno de los sujetos. La significación estadística se obtiene aplicando la  $t$  de Student para datos relacionados.

Según Plewis (1985), los modelos de la regresión son más apropiados para la medida del cambio que los modelos incondicionales, puesto que los primeros permiten tener en cuenta la dirección temporal, es decir, la direccionalidad del cambio.

### **3.2.3. Técnicas de análisis del diseño longitudinal de múltiples observaciones**

El objetivo principal de los análisis estadísticos aplicados a los diseños 1GMO, tanto del AVAR mixto como de los análisis alternativos, es obtener una curva de crecimiento que represente cómo la media poblacional cambia con el tiempo (Bock, 1979; Rogosa, Brandt y Zimowski, 1982; Stoolmiller, 1995).

A fin de determinar el tipo de tendencia o presencia de alguna orientación sistemática de las medias en función de las ocasiones de

observación (ajuste de curvas de crecimiento), se requiere la estimación de los coeficientes polinómicos ortogonales.

### A) Modelo de AVAR mixto

El modelo estadístico aplicable al diseño 1GMO es el AVAR mixto, con variables fijas y aleatorias (Arnau, 1990b, 1995e, 1997; Bock, 1975; Kirk, 1982; Pascual, 1995; Riba, 1987, 1990; Winer, 1971), cuya expresión viene dada por la siguiente ecuación:

$$Y_{ij} = \mu + \eta_i + \alpha_j + \varepsilon_{ij} \quad (19)$$

donde

$Y_{ij}$  = puntuación del  $i$ ésimo sujeto bajo la ocasión  $j$ .

$\mu$  = media común a todas las observaciones.

$\eta_i$  = componente específico asociado al sujeto  $i$  (constante a lo largo de todas las observaciones).

$\alpha_j$  = efecto general de la  $j$ ésima ocasión, común a todos los sujetos.

$\varepsilon_{ij}$  = componente de error asociado al sujeto  $i$  bajo la ocasión  $j$ .

### ESPECIFICACIONES DEL MODELO:

1) El factor sujeto es una variable aleatoria independientemente distribuida:  $\eta_i \sim NID(0, \sigma_\eta^2)$ .

2)  $\varepsilon_{ij} \sim NID(0, \sigma_\varepsilon^2)$

3) Se asume que  $\varepsilon_{ij}$  es independiente de  $\eta_i$ .

4) Los niveles del factor ocasiones de observación son fijos:

$$\sum_{j=1}^I \alpha_j = 0$$

5) Supuesto de uniformidad o de simetría combinada: homogeneidad de las variancias y covariancias de la población.

6) Supuesto de esfericidad o circularidad: igualdad de las variancias de las diferencias entre todos los pares de medidas repetidas (Huynh y Feldt, 1970; Rouanet y Lepine, 1970). Por ejemplo, para una matriz 3x3, debe cumplirse que  $\sigma_D 12^2 = \sigma_D 13^2 = \sigma_D 23^2$  donde  $\sigma_{Dij}^2 = \sigma_i^2 + \sigma_j^2 - 2\sigma_{ij}$  (Games, 1990; Scheirs, 1992). Esta condición es suficiente para la prueba de hipótesis mediante la  $F$  normal.

Con el propósito de verificar el presupuesto de uniformidad se aplica la prueba de Box (1950), y con objeto de probar el presupuesto de circularidad es de interés la prueba de Mauchley (1940).

### **B) Alternativas al análisis del modelo mixto**

Cuando se cumplen los supuestos de uniformidad o esfericidad de la matriz de covariancia del diseño, la prueba de hipótesis más adecuada y potente es el AVAR (Arnau, 1990b, 1995e, 1997; Riba, 1987, 1990). En caso de no cumplirse los supuestos del modelo mixto, se obtiene un razón  $F$  positivamente sesgada, afectando tanto al error Tipo I como a la potencia de la prueba (Pascual, 1995). Este es el motivo por el que existen tres alternativas de análisis (figura 3.3): a) modelo de AVAR, pero simplificando los grados de libertad para hallar la  $F$  teórica (prueba conservadora); b) modelo de AVAR corrigiendo los grados de libertad de la razón  $F$  mediante la  $\varepsilon$  de Greenhouse y Geisser (1959), y c) AMVAR.

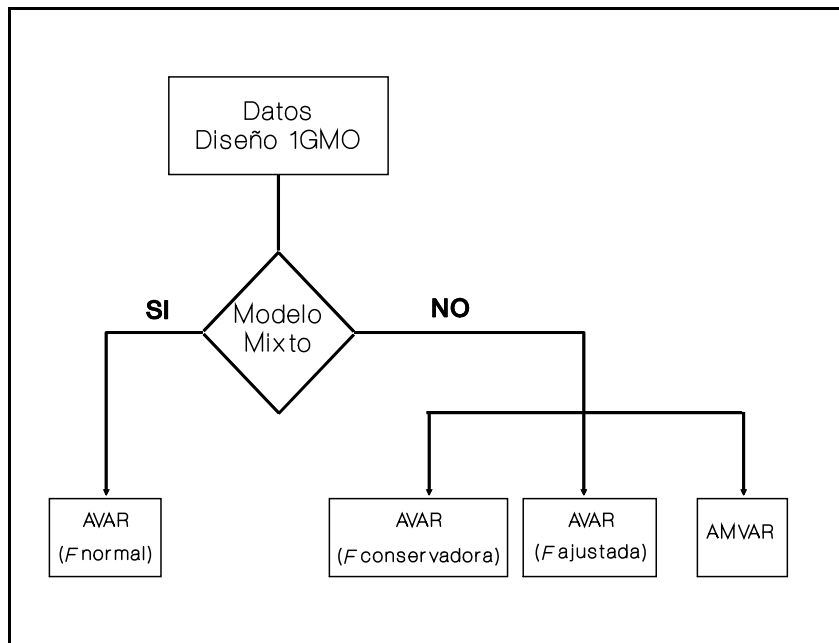


Figura 3.3. Análisis de datos del diseño 1GMO.

### 3.2.4. Técnicas de análisis del diseño split-plot

Desde un punto de vista estadístico, el diseño split-plot recibe el nombre de diseño mixto (Pardo y San Martín, 1994; Pascual, 1995) o diseño de perfiles (Arnau, 1995f; Greenhouse y Geisser, 1959; Morrison, 1967).

El diseño split-plot combina, en una misma estructura, el efecto de la dimensión temporal y una variable pronóstica capaz de afectar a las respuestas de los sujetos.

#### A) Análisis de perfiles

El análisis que se aplica a los diseños split-plot recibe el nombre genérico de análisis de perfiles. A través de este análisis, se formulan -de manera ordenada- las siguientes hipótesis (Arnau, 1995f):

- 1) La primera hipótesis responde a si los perfiles de los grupos de la variable de clasificación son paralelos. En caso de rechazarse la hipótesis nula, se concluye que el perfil de cada grupo depende de las ocasiones de observación (interacción significativa entre grupo y ocasión de observación) y, por lo tanto, es innecesario el planteamiento de cualquier otra hipótesis.
- 2) En el supuesto de que los perfiles sean paralelos, la segunda hipótesis hace referencia a si dichos perfiles son coincidentes. Es decir, se plantea la existencia de diferencias entre los grupos (variable de clasificación significativa).
- 3) En caso de que los perfiles sean coincidentes, la siguiente hipótesis que se postula es si los perfiles son constantes. Esta última hipótesis permite probar si existe algún cambio o tendencia que sea función del tiempo (variable ocasión de observación significativa).

## B) Modelo de AVAR mixto

El modelo estructural del diseño split-plot (ecuación 20) puede considerarse como una extensión del modelo de medidas repetidas de un solo grupo y múltiples observaciones.

$$Y_{ijk} = \mu + \alpha_j + \eta_{ij} + \beta_k + (\alpha\beta)_{jk} + (\eta\beta)_{ikj} + \varepsilon_{ijk} \quad (20)$$

donde

$Y_{ijk}$  = puntuación del sujeto  $i$  dentro del grupo  $j$  y en la ocasión  $k$ .

$\mu$  = media común a todas las observaciones.

$\alpha_j$  = efecto del  $j$ ésimo nivel de la variable de clasificación o grupo.

$\eta_{ij}$  = efecto asociado al  $i$ ésimo sujeto del grupo  $j$  (componente de error entre).

$\beta_k$  = efecto de la ocasión de observación  $k$ .

$(\alpha\beta)_{jk}$  = interacción entre el grupo  $j$  y la ocasión de observación  $k$ .

$(\eta\beta)_{ikj}$  = interacción entre el sujeto  $i$  y la ocasión  $k$  dentro del grupo de clasificación  $j$  (componente de error intra).

$\varepsilon_{ijk}$  = componente de error o residual.

De acuerdo con esta estructura hay dos términos de error: el error *entre* contra el que se contrasta la variable de grupos, y el error *intra* contra el que se contrasta el efecto de la variable ocasiones y la interacción grupos x ocasiones.

#### ESPECIFICACIONES DEL MODELO:

- 1) El componente de error es una variable aleatoria con distribución normal e independiente en todos los grupos:

$$\varepsilon_{ijk} \sim NID(0, \sigma_{\varepsilon}^2)$$

- 2) Con respecto al componente de sujetos se postula que:

$$\eta_{ij} \sim NID(0, \sigma_{\eta}^2)$$

$$(\eta\beta)_{ikj} \sim NID(0, \sigma_{\eta\beta}^2)$$

- 3) Supuesto de simetría combinada para las variables de medidas repetidas (ocasión de observación e interacción grupo de clasificación x ocasión).

### 3.3. Diseños de cohortes (DC)

El estudio del cambio en el proceso de desarrollo es de gran interés dentro del ámbito de la metodología longitudinal. La principal cuestión que se plantea es la contribución de los efectos de edad, período (o tiempo de medida) y cohorte en la explicación del cambio ontogénico o madurativo. En este sentido, el análisis de cohortes se orienta a la identificación de cada uno de estos componentes.

Los DC pueden definirse como estructuras de investigación donde los datos están organizados en función de distintas cohortes, edades y períodos, con el propósito de estudiar el efecto de cada factor y su posible interacción. Los conceptos relacionados con esta clase de estudios son los siguientes.

*Cohorte.* Grupo o agregado de individuos, dentro de una población

específica, que ha experimentado el mismo acontecimiento vital en un intervalo de tiempo determinado (Glenn, 1977; Ryder, 1965; Visser, 1985). Aunque la variable más usada para delimitar una cohorte es la fecha de nacimiento, puede darse un número distinto de variables que definen cada cohorte (Glenn, 1977; Markus, 1985; Menard, 1991; Ryder, 1965; Wall y Williams, 1970). Por ejemplo, los individuos casados en un período establecido de tiempo, los sujetos que asistieron a la escuela en un determinado año, etc.

*Efectos de edad.* Movimientos a largo plazo asociados al proceso de desarrollo, a través del ciclo vital. En otras palabras, el efecto de edad es un indicador de toda clase de procesos y hechos que están asociados al crecimiento.

*Efectos de período.* Fluctuaciones de los datos debidas a hechos específicos o circunstancias que ocurren en determinados puntos de tiempo. Por lo tanto, el efecto de período puede ser considerado como un indicador de los impactos de un conjunto de hechos que ocurren entre los momentos de observación, así como de las influencias de procesos a más largo plazo (por ejemplo, cambios en programas educativos, industrialización, etc.).

*Efectos de cohortes.* Distinciones intercohortes duraderas, atribuidas a una impronta o característica común de los miembros de cada cohorte.

### **3.3.1. Características del DC**

- 1) Los DC están directamente asociados a los procesos de cambio del individuo a lo largo del tiempo, y a las diferencias individuales y de grupo.
- 2) En los DC se analizan muestras de miembros representativos de las diferentes cohortes de población.
- 3) Tradicionalmente, las estrategias de análisis utilizadas para el estudio de los procesos del desarrollo han sido los diseños transversales



(múltiples muestras de distintas edades se observan en un mismo período), los diseños longitudinales (una misma muestra es observada a través de las edades y períodos) y los diseños de retardo temporal (distintas muestras de cohortes son observadas a una misma edad y con una sola observación). En la figura 3.4 se representan estas tres estrategias de análisis.

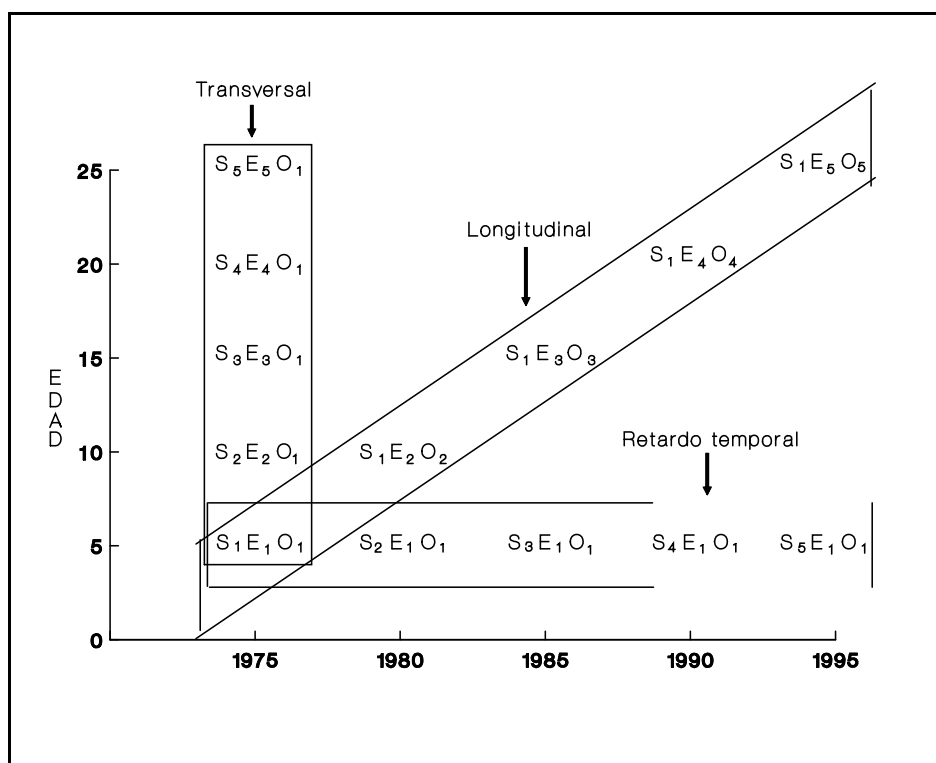


Figura 3.4. Representación gráfica de los diseños tradicionales en el estudio del desarrollo, donde S<sub>1</sub>-S<sub>5</sub> simbolizan muestras distintas de sujetos; E<sub>1</sub>-E<sub>5</sub>, diferentes edades, y O<sub>1</sub>-O<sub>5</sub>, las diversas observaciones (modificada de Baltes *et al.*, 1988; p. 122).

- 4) Cada uno de los diseños representados en la figura 3.4 generan patrones de confundido entre los factores que afectan al desarrollo: edad, cohorte y período (Hagenaars, 1990; Schaie, 1965, 1970, 1972, 1992).

### 3.3.2. Modelo general del desarrollo

Cuando al estudio del cambio psicológico o evolutivo se aplican diseños transversales, se confunde el efecto de la variable edad con el efecto de la variable de cohortes. Por otra parte, cuando se utilizan diseños longitudinales, la confusión se produce entre el efecto del período y la edad. Por último, con diseños de retardo temporal, suele confundirse el efecto de la variable período con el efecto de la variable cohortes, y el efecto de la variable edad y período de observación.

Ante estas dificultades, Schaie (1965) propuso un modelo general o trifactorial del desarrollo y derivó del mismo una serie de diseños de investigación. De acuerdo con este modelo, se tiene la siguiente expresión:

$$D = f(E, C, P) \quad (21)$$

donde  $D$ , que simboliza alguna dimensión de cambio o desarrollo, es función de la edad cronológica ( $E$ ), la cohorte ( $C$ ) y el período o tiempo de observación ( $P$ ).

A partir del modelo de la ecuación (21) se deriva el *diseño longitudinal mixto*, que se divide en tres modelos de *diseños secuenciales* (transverso, de tiempo y de cohorte). En cualquiera de las tres estrategias propuestas por Schaie, sigue persistiendo la confusión de los efectos de edad, cohorte y período. Sin embargo, la estimación independiente del efecto de cada variable puede lograrse mediante la aplicación, por separado, de los diseños secuenciales a un conjunto de datos (Schaie, 1974). Para ello, se disponen los datos en términos de dos variables. De este modo, se tiene los siguientes formatos de diseño, que se analizan separadamente.

En el formato del diseño secuencial transverso se cruza la variable cohorte con la variable tiempo de medida o período, sin tener en cuenta el factor edad (tabla 3.1).

Tabla 3.1. Diseño secuencial transverso: matriz de datos para un análisis de cohorte x tiempo de medida.

Cohorte (Edad de nacimiento)	Tiempo de medida		
	1970	1980	1990
1940	Celda 1	Celda 2	Celda 3
1930	Celda 4	Celda 5	Celda 6
1920	Celda 7	Celda 8	Celda 9

El análisis que podría aplicarse a los datos de la tabla 3.1, bajo el supuesto que los sujetos de cada cohorte repiten medida a lo largo de los tres períodos, es un AVAR mixto con una variable entre (cohortes) y una variable intra (tiempo de medida). El análisis termina en este primer diseño cuando se parte del supuesto que el efecto de la edad es nulo y, por lo tanto, no interesa en el estudio. En caso contrario, se prosigue con el resto de análisis.

En una segunda fase, se lleva a cabo el diseño secuencial de tiempo, que permite estudiar el efecto de la edad, así como la interacción edad x período, a partir de una matriz de datos con el formato de la tabla 3.2.

Tabla 3.2. Diseño secuencial de tiempo: matriz de datos para un análisis de edad x tiempo de medida (modificada de Williams, 1991; p. 647).

Edad en la medida	Tiempo de medida		
	1970	1980	1990
30	Celda 1		
40	Celda 4	Celda 2	
50	Celda 7	Celda 5	Celda 3
60		Celda 8	Celda 6
70			Celda 9

El problema fundamental del diseño secuencial de tiempo son las celdas vacías. La solución que propuso Schaie (1965, 1977), entre otros, consiste en utilizar únicamente las celdas 4, 2, 7 y 5, o bien las casillas 5, 3, 8 y 6, aplicando un AVAR mixto con una variable entre (edad) y otra intra (tiempo de medida).

Por último, se completa el análisis a través del diseño secuencial de cohorte, cuya matriz de datos se representa en la tabla 3.3. En esa tercera parte del análisis, es importante centrarse en el estudio de la interacción cohorte x edad, puesto que los efectos de cohorte y edad ya han sido estimados con anterioridad (tablas 3.1 y 3.2, respectivamente).

Tabla 3.3. Diseño secuencial de cohorte: matriz de datos para un análisis de cohorte x edad (modificada de Williams, 1991; p. 643).

Cohorte	Edad en el tiempo de medida				
Edad de nacimiento	30	40	50	60	70
1940	Celda 1	Celda 2	Celda 3		
1930		Celda 4	Celda 5	Celda 6	
1920			Celda 7	Celda 8	Celda 9

La tabla 3.3 muestra también seis celdas vacías, de manera que puede aplicarse dos análisis de la variancia paralelos: el primero con las celdas 2-5, y el segundo con las celdas 5-8.

En resumen, los diseños secuenciales permiten estudiar los efectos principales de edad, cohorte y período, así como sus interacciones. Sin embargo, la interpretación continua siendo difícil, debido a la existencia de dependencia lineal entre las variables que intervienen (*problema de identificación*). En otras palabras, debido a la imposibilidad de separar los efectos de cohortes, edades y períodos en los modelos lineales empleados (Adam, 1978; Arnau, 1997; Horn y Donaldson, 1977). Además, desde el punto de vista analítico y de cálculo, se añade el problema de la falta de algunas celdas en la segunda y tercera parte del análisis. Con objeto de solventar estos inconvenientes, los modelos secuenciales requieren asumir los siguientes presupuestos: a) efecto nulo de la variable edad en el análisis cohorte x período; b) efecto nulo de la variable de cohorte en el análisis edad x período, y c) en el diseño cohorte x edad se asume el no efecto de la variable período.

### 3.4.3. Modelos de análisis

El objetivo principal de los estudios del desarrollo es estimar, por separado, la contribución de los efectos de edad, período y cohorte sobre una variable de respuesta de los sujetos pertenecientes a cohortes distintas. Sin embargo, aunque desde el punto de vista analítico es posible detectar el efecto de los distintos factores, su interpretación es difícil debido al problema de la identificación que ha llevado a diversas controversias (Adam, 1978; Costa y McCrae, 1982; Donaldson y Horn, 1992, entre otros). Siguiendo la solución del análisis por separado de cada uno de los diseños, los diferentes modelos estructurales corresponden a un AVAR mixto en los diseños secuencial transverso y secuencial de tiempo, y a un AVAR en el diseño secuencial de cohorte.

El modelo estructural del AVAR mixto correspondiente al diseño secuencial transverso viene expresado de la siguiente forma:

$$Y_{ijk} = \mu + \alpha_j + \eta_{ij} + \beta_k + (\alpha\beta)_{jk} + (\eta\beta)_{ikj} + \varepsilon_{ijk} \quad (22)$$

donde

$Y_{ijk}$  = puntuación del sujeto  $i$  dentro de la  $j$ ésima cohorte y en el  $k$ ésimo período.

$\mu$  = media total.

$\alpha_j$  = efecto debido a la  $j$ ésima cohorte.

$\eta_{ij}$  = efecto asociado al  $i$ ésimo sujeto de la cohorte  $j$  (componente de error entre).

$\beta_k$  = efecto del  $k$ ésimo período.

$(\alpha\beta)_{jk}$  = efecto de la interacción entre la  $j$ ésima cohorte y el  $k$ ésimo período.

$(\eta\beta)_{ikj}$  = interacción entre el  $i$ ésimo sujeto y el período  $k$  dentro de la  $j$ ésima cohorte (componente de error intra).

$\varepsilon_{ijk}$  = error residual.

En el diseño secuencial de tiempo, el modelo estructural del AVAR mixto es igual al de la ecuación (22), con la diferencia que el subíndice  $j$  corresponde a la edad. Por último, el diseño secuencial de cohorte se resuelve mediante la aplicación del AVAR, cuyo modelo estructural es el que se detalla a continuación:

$$Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk} \quad (23)$$

donde

$Y_{ijk}$  = puntuación del  $i$ ésimo sujeto dentro de la  $j$ ésima cohorte y en la  $k$ ésima edad.

$\mu$  = media total.

$\alpha_j$  = efecto debido a la  $j$ ésima cohorte.

$\beta_k$  = efecto de la  $k$ ésima edad.

$(\alpha\beta)_{jk}$  = interacción entre la  $j$ ésima cohorte y la  $k$ ésima edad.

$\varepsilon_{ijk}$  = error residual.

Las especificaciones del modelo del AVAR de carácter mixto y el AVAR de dos factores son las ya expuestas a propósito del modelo estructural del AVAR mixto en el diseño split-plot y del AVAR de las puntuaciones después en el DGCNE, respectivamente.

### 3.4. Diseños en panel (DP)

El *diseño longitudinal en panel* requiere que a partir de una misma muestra de sujetos se tomen registros u observaciones de dos o más variables, en intervalos de tiempo de bastante amplitud o tandas (*waves*). Puesto que las medidas se toman en tandas o cortes temporales a lo largo de una serie de puntos discretos en el tiempo, los DP constituyen una combinación de los diseños de series temporales y los diseños transversales (Arnau, 1997; Hsiao, 1986; Rogosa, 1979).

La estructura del DP admite diferentes modalidades, según se combinen la cantidad de tandas y variables (tabla 3.4). La variedad más simple es el DP de dos tandas y dos variables o diseño 2W2V (Kenny, 1975b). La extensión de este formato es el diseño de tres tandas y dos variables (3W2V). Además, el formato del DP también puede ampliarse con respecto a la cantidad de variables.

Tabla 3.4. Clasificación del diseño longitudinal en panel (Arnau, 1997).

TANDAS	VARIABLES	MODELO
Dos tandas	Dos variables	2W2V
Tres tandas	Dos variables	3W2V
.....	.....	.....
<i>N</i> tandas	<i>N</i> variables	NWNV

### 3.4.1. Características del DP

- 1) Amplia tradición en investigación social.
- 2) El diseño longitudinal en panel utiliza una única muestra de sujetos.
- 3) Los puntos de observación del diseño se denominan tandas y, en cada una de ellas, se pueden tomar medidas de dos o más variables de forma simultánea.
- 4) Instrumento eficaz para el estudio del cambio, así como de las relaciones de causa-efecto que se dan, a través del tiempo, entre las variables.
- 5) La representación gráfica del diseño 2W2V se muestra en la figura 3.5. A partir de esta estructura, se calculan las posibles correlaciones, representadas por las rectas correspondientes.



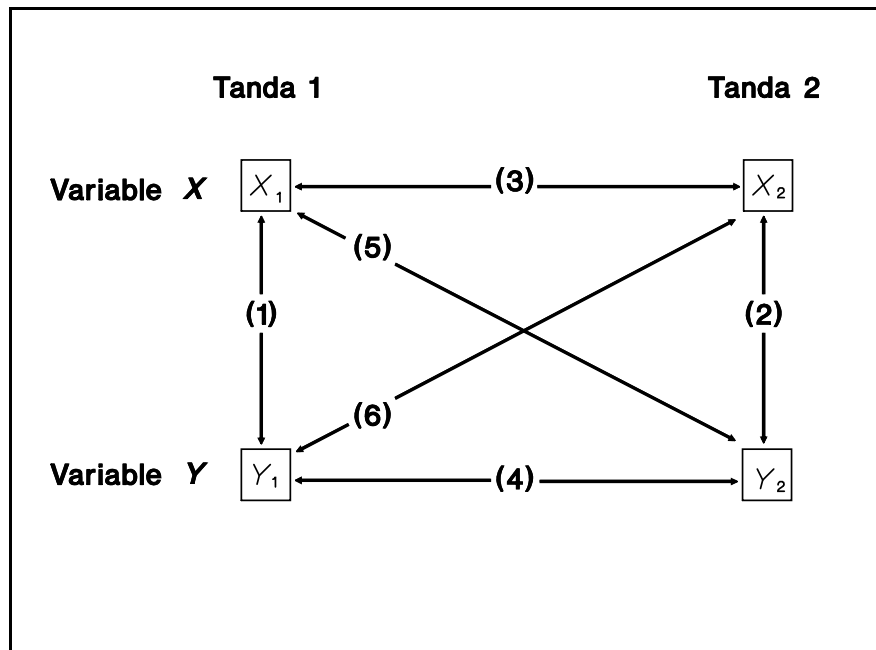


Figura 3.5. Representación de las correlaciones del diseño 2W2V: (1)  $r_{X1Y1}$ ; (2)  $r_{X2Y2}$ ; (3)  $r_{X1X2}$ ; (4)  $r_{Y1Y2}$ ; (5)  $r_{X1Y2}$ , y (6)  $r_{Y1X2}$  (modificada de Rogosa, 1980b; p. 149).

- 6) La causalidad entre las variables se apoya en el retardo temporal de una sobre la otra.

### 3.4.2. Correlación en panel de retardos cruzados

En la figura 3.5 están representados los tres pares de correlaciones (Arnau y Gómez, 1995; Kenny, 1975b, 1979): a) *autocorrelaciones* ( $r_{X1X2}$  y  $r_{Y1Y2}$ ); b) *correlaciones sincrónicas* ( $r_{X1Y1}$  y  $r_{X2Y2}$ ), y c) *correlaciones de retardos cruzados* ( $r_{X1Y2}$  y  $r_{Y1X2}$ ). Las autocorrelaciones son un buen indicador de la estabilidad de la medida a lo largo del tiempo; las correlaciones denominadas sincrónicas reflejan las asociaciones contemporáneas entre las variables X e Y en dos puntos del tiempo o tandas, y las correlaciones de retardos cruzados constituyen la asociación temporal entre X e Y. Como es obvio, estas últimas correlaciones intervienen en la verificación de la causalidad.

La lógica del análisis de correlaciones de retardos cruzados se fundamenta en el concepto de *retardo temporal*, presente cuando se asume una relación causal entre dos variables. En este sentido, en el caso que por ejemplo X cause a Y, "el estado presente de X debe estar más fuertemente asociado al estado futuro de Y que a su estado pasado o presente" (Arnau y Gómez, 1995; pp. 343-344).

En función del signo que toma la diferencia entre las correlaciones  $r_{X_1Y_2}$  y  $r_{Y_1X_2}$ , se infiere la causalidad entre X e Y (Campbell, 1963). Cuando la diferencia entre estas dos correlaciones es positiva ( $r_{X_1Y_2} > r_{Y_1X_2}$ ), se desprende una relación causal de manera que X afecta a Y ( $X \rightarrow Y$ ). Por el contrario, cuando el signo es negativo, la causalidad es inversa ( $Y \rightarrow X$ ). Antes de inferir una relación causal, sería conveniente descartar cualquier explicación de acuerdo con algún modelo espurio (Arnau y Gómez, 1995; Kenny, 1973, 1975b, 1979), es decir, que la causa de la correlación se deba a una tercera variable no tenida en cuenta.

La hipótesis de igualdad de las correlaciones cruzadas o hipótesis de espureidad (ecuación 24) puede ponerse a prueba mediante la fórmula de Pearson-Filon (Kenny, 1975b, 1979; Peters y Van Voorhis, 1940) o transformación z (ecuación 25), cuya distribución es normal.

$$H_0: r_{X_1Y_2} - r_{Y_1X_2} = 0 \quad (24)$$

Sean las variables  $X_1 = 1$ ,  $Y_1 = 2$ ,  $X_2 = 3$  y  $Y_2 = 4$ , se tiene que para la prueba de espureidad:

$$z = \frac{N^{1/2} (r_{14} - r_{23})}{[(1 - r_{14}^2)^2 + (1 - r_{23}^2)^2 - k]^{1/2}} \quad (25)$$

donde

$N$  = tamaño muestral.

$$k = (r_{12} - r_{24}r_{14})(r_{34} - r_{24}r_{23}) + (r_{13} - r_{12}r_{23})(r_{24} - r_{12}r_{14}) + (r_{12} - r_{13}r_{23})(r_{34} - r_{13}r_{14}) + (r_{13} - r_{14}r_{34})(r_{24} - r_{34}r_{23})$$

Entrando en las tablas de z, se halla la probabilidad asociada a dicho

valor y se comprueba la significación estadística. En caso de rechazarse la prueba de espureidad, se infiere una relación causal. Sin embargo, para la inferencia de la causalidad, no es suficiente que la diferencia entre las correlaciones de retardos cruzados sea estadísticamente significativa, sino que además deben cumplirse dos asunciones previas (Kenny, 1975b, 1979):

- 1) *Condición de estacionariedad* (ecuación 26): igualdad de las correlaciones sincrónicas a lo largo del tiempo, a fin de probar la estabilidad del sistema.

$$H_0: r_{X1Y1} - r_{X2Y2} = 0 \quad (26)$$

Esta hipótesis de nulidad se prueba estadísticamente a través del estadístico de la fórmula de Pearson-Filon (ecuación 25).

- 2) *Condición de sincronía*: las dos variables han de ser medidas en el mismo punto del tiempo.

El análisis de datos de los DP también puede resolverse mediante modelos de la regresión o de ecuaciones estructurales. Estas dos técnicas permiten una modelación de los efectos causales más adecuada que el análisis de correlaciones de retardos cruzados. De este modo, es posible estimar los parámetros de un modelo causal. Además, la técnica de los modelos de ecuaciones estructurales permite el planteamiento de modelos alternativos.

## 4. MODELADO ESTADÍSTICO

### 4.1. Modelos estadísticos

Los diseños cuasi-experimentales y diseños longitudinales expuestos representan estrategias particulares de obtención de datos que servirán para la posterior validación de algún modelo, propuesto como la representación de una realidad concreta.

Los modelos matemáticos pueden ser deterministas o probabilistas

(Neter *et al.*, 1985). Estos últimos también son denominados estocásticos. Un *modelo determinista* se define como aquel cuyas variables predictoras determinan por completo a la variable criterio modelada. Por otra parte, un *modelo probabilista* presupone la existencia de un término de error que dé cuenta de la desviación entre la realidad observada (empírica) y la representada teóricamente por el modelo. En ciencias del comportamiento, los modelos tienen carácter probabilista, de manera que pueden representarse por la siguiente expresión general (Judd y McClelland, 1989; Judd, McClelland y Culhane, 1995):

$$\text{DATOS} = \text{MODELO} + \text{ERROR}$$

o como propone Tukey (1991):

$$\text{DATOS} = \text{AJUSTE} + \text{RESIDUAL}$$

Siguiendo una formulación estrictamente matemática, los distintos modelos suelen expresarse por la siguiente ecuación general de carácter lineal:

$$Y = f(x) + g(\varepsilon) \quad (27)$$

donde  $Y$  representa la variable de respuesta u observada del modelo que constituye una combinación aditiva de alguna función de variables fijas  $f(X)$  101 y otra función de componentes aleatorios  $g(\varepsilon)$ . 102 Por tanto, de la ecuación (27) se deriva que cualquier observación es la suma de dos componentes: uno sistemático o determinista y otro aleatorio desconocido. Los parámetros asociados a la parte determinista del modelo son de gran interés, ya que representan la magnitud de cambio o el efecto causal de las variables, de acuerdo con el tipo de estrategia de recogida de datos empleada.

Resulta evidente que el objetivo de todo investigador es proponer y hallar el modelo que mejor se ajuste a los datos observados y, en consecuencia, que tenga un componente de error mínimo. En el caso extremo de residual cero, los datos estarían perfectamente representados por el modelo investigado. Al proceso que persigue este fin se le denomina modelado estadístico.

## 4.2. Fases del modelado estadístico

El procedimiento de modelado estadístico se compone de una serie de fases que definen un proceso iterativo en la búsqueda del modelo más adecuado. La validación del modelo teórico requiere su traducción en una ecuación matemático-estadística, que se contrastará de nuevo con los datos obtenidos. Este proceso de modelación consta de las siguientes etapas:

- A) Selección de un modelo estadístico adecuado.
- B) Estimación de los parámetros.
- C) Evaluación de la adecuación del modelo propuesto a los datos.
- D) Interpretación de los resultados.

### A) Selección

La elección de un modelo estadístico adecuado para la representación de los datos es la primera fase del modelado. Dependiendo de la forma del componente sistemático, es decir, del modo como se relacionan parámetros y variables independientes, los modelos estadísticos que comparten la estructura mostrada en la ecuación (27) pueden ser (Seber, 1977, 1980): a) modelos lineales tanto en las variables como en los parámetros; b) modelos lineales en los parámetros y no en las variables, y c) modelos no lineales respecto a las variables y a los parámetros. Los dos primeros tipos quedan englobados en el Modelo Lineal General (MLG), y el tercero responde a los modelos no lineales.

En función de la naturaleza de las variables independientes, los modelos lineales se subdividen en: a) modelos de regresión, cuando las variables de la parte determinista son cuantitativas; b) modelos de diseño experimental o de análisis de la variancia, cuando las variables independientes son de naturaleza categórica, y c) modelos con covariables o de análisis de la covariancia, cuando están presentes tanto variables de tipo cuantitativo como categórico. Por tanto, en el primer paso del proceso de modelado estadístico, el investigador deberá seleccionar uno de estos tres modelos generales. Esta fase se acompaña de dos etapas que deben estar en mutua correspondencia (Riba, 1987, 1990): la

especificación y la identificación.

La *especificación* consiste en establecer la ecuación estructural, es decir, se determinan todos los elementos del modelo (parámetros y variables independientes). La *identificación* permite validar o modificar el modelo, en función de que se disponga o no de unidades de información (datos observables) suficientes para proceder a la estimación de sus parámetros. En función de ambas cantidades, los modelos pueden ser: a) subidentificados, cuando se dispone de menos información que de parámetros totales del modelo; b) sobreidentificados, cuando se poseen más unidades de información que parámetros, y c) identificados, cuando se dispone de las mismas unidades de información que de parámetros. La condición necesaria para que un modelo sea identificable es que el número de parámetros a estimar sea inferior o igual al número de unidades de información disponible (Riba, 1987, 1990).

Del proceso de selección puede resultar un modelo no adecuado a los datos, como consecuencia, principalmente, de alguno de los siguientes *errores de especificación* (Pedhazur 1982): a) definición de un modelo lineal cuando es más apropiado un modelo no lineal; b) omisión de alguna variable independiente relevante, y c) inclusión de alguna variable explicativa irrelevante.

## **B) Estimación**

El criterio usualmente utilizado en la estimación de los parámetros de un modelo lineal es el de mínimos cuadrados, cuyo objeto es la obtención de un conjunto de estimadores paramétricos que minimicen el componente aleatorio del modelo, es decir, la suma de cuadrados de los residuales o diferencias entre los valores observados y los pronosticados por el componente sistemático.

## **C) Evaluación**

Una vez estimados los parámetros, se evalúa la adecuación del modelo como sistema de representación de los datos. Esta tercera fase incluye la

verificación de los supuestos del modelo y la valoración de la bondad de ajuste del modelo a los datos de acuerdo con un criterio estadístico.

El ajuste del modelo puede proponerse desde una estrategia global o parcial. En el primer caso, se pretende detectar si el modelo tiene una capacidad de representación de los datos mayor a la esperada por azar. Una vez realizado el ajuste global del modelo, una segunda estrategia de ajuste parcial intenta detectar a qué componentes se debe su adecuación, con el fin de mejorar el modelo, eliminando aquellos elementos que no contribuyen de forma significativa al objeto perseguido (Judd y McClelland, 1989; Maxwell y Delaney, 1990). En otras palabras, con el proceso de ajuste parcial se pretende conseguir un modelo óptimo o estructura en el que todos los elementos sean relevantes (Seber, 1977).

En resumen, en la fase de evaluación se comprueba la consistencia matemática del modelo y el ajuste entre los resultados derivados del modelo y los obtenidos empíricamente de las variables estudiadas.

Las tres etapas descritas hasta aquí -selección, estimación y evaluación- deben desarrollarse en un proceso iterativo, cuyo objetivo es la búsqueda del mejor ajuste posible, de tal manera que con el menor número de parámetros se consiga minimizar al máximo el componente de error del modelo.

#### **D) Interpretación**

La última fase del modelado estadístico consiste en la interpretación de los parámetros del modelo estimado. En esta etapa deberá justificarse la conveniencia del modelo, tanto desde el punto de vista estadístico como teórico. Desde su referente estadístico, debe probarse que el modelo propuesto es el más adecuado estadísticamente y, desde su referente teórico, debe proporcionarse la integración del conocimiento adquirido en el entramado de la teoría desde la que fue propuesto. En caso negativo, se desecha el modelo inicialmente considerado, lo que obligará al investigador a retroceder de nuevo hasta el nivel teórico, en busca de otro modelo para explicar los datos disponibles.

### 4.3. Inferencia estadística: contraste de hipótesis

Los parámetros asociados al componente determinista representan la magnitud de cambio o el efecto causal de las variables. De ahí, que el interés del investigador se centra en hallar el nivel de significación de los parámetros, lo cual se consigue mediante dos estrategias: probando individualmente los parámetros con el estadístico  $t$ , o bien probando conjuntamente la totalidad de ellos con el estadístico  $F$ . En ambos casos, se prueba la hipótesis nula ( $H_0$ ) con respecto a cada parámetro o al conjunto de ellos, respectivamente.

Antes de proseguir con los aspectos más relevantes del contraste de hipótesis, tales como los errores de Tipo I y Tipo II, consideramos de interés presentar un breve desarrollo histórico. Tradicionalmente existe una controversia entre dos corrientes. La escuela de Fisher considera las pruebas estadísticas como pruebas de significación (Fisher, 1935, 1950, 1955), mientras que la escuela de Neyman-Pearson las conceptúa como pruebas de decisión (Neyman y Pearson, 1928, 1933, 1936).

Según Chase y Tucker (1976), esta controversia entre tradiciones puede reducirse a que ante una diferencia significativa, un defensor del enfoque fisheriano concluiría que la  $H_0$  no es válida (con lo que se prueba la existencia del fenómeno que se esté estudiando); mientras que un investigador de la tradición de Neyman-Pearson, simplemente rechazaría la  $H_0$  para esta ocasión particular.

Otra distinción importante entre estas dos aproximaciones se refiere a si el nivel de significación ( $\alpha$ ) empleado en el análisis debe fijarse anticipadamente. Aunque Fisher (1935, 1950, 1955) daba prioridad a un nivel de significación de 0.05, nunca prescribió que tal nivel debiera mantenerse fijo o que debiera establecerse antes de llevar a cabo el experimento. Por contra, la posición de Neyman-Pearson (1928, 1933, 1936) requiere que el nivel de significación se determine antes de cualquier análisis estadístico y que el investigador se adhiera a él para todas las decisiones estadísticas. En este sentido, si el nivel de significación no se estableciera con anterioridad, se podría caer en el error de fijar un valor  $\alpha$  basado más en los resultados que en



una estimación de la probabilidad.

Una tercera divergencia entre las dos teorías se refiere a la interpretación de resultados positivos y negativos de la investigación experimental. De acuerdo con Fisher (1935, 1950, 1955), se puede afirmar que el efecto no es cero cuando se rechaza la  $H_0$ , pero no es posible concluir que sea cero cuando se acepta. Lo que realmente significa la afirmación falsa de que la  $H_0$  es cierta es que existe un efecto despreciable o trivial (Cohen, 1988). Por el contrario, el planteamiento de Neyman-Pearson (1928, 1933, 1936) postula la existencia de una hipótesis alternativa ( $H_1$ ) exacta del tamaño del efecto. Esta proposición llevó a Neyman y Pearson (1928, 1933) al concepto de error de Tipo II (aceptar la  $H_0$  cuando es falsa), relacionado con el de potencia (Bono y Arnau, 1995). En cuanto al error de Tipo I, característico de la tradición fisheriana, cabe destacar que no consiste en rechazar la  $H_0$  cuando es verdadera sino en el nivel de riesgo que el investigador está dispuesto a aceptar cuando interpreta los resultados de una prueba estadística.

Los manuales de psicología y educación transmitieron, primero, el mensaje fisheriano y después de la II Guerra Mundial, los autores comenzaron a darse cuenta del impacto de la aproximación de Neyman-Pearson y a sustituir la teoría anterior. Sin embargo, los conceptos de error Tipo II y potencia fueron incluidos por la mayoría de investigadores, dentro del marco fisheriano. El resultado fue una teoría híbrida con ideas contrapuestas (prueba de la  $H_0$ , siguiendo a Fisher, y error de Tipo II y especificación del error de Tipo I antes de la recogida de datos, siguiendo a Neyman y Pearson), lo que conllevó una confusión del significado de conceptos básicos (Bakan, 1966; Oakes, 1986) y, en algún grado, una explicación del continuo descuido acerca de la potencia (Bono y Arnau, 1995). Esta estadística híbrida se enseña, normalmente, como estadística *per se*, sin mención de los originadores de las respectivas ideas (Gigerenzer y Murray, 1987).

### 4.3.1. Errores de Tipo I y Tipo II

Es sabido que al tomar una decisión estadística (rechazar o aceptar la  $H_0$ ) se corre el riesgo de cometer un error, debido a la dificultad de discriminar con certeza entre las hipótesis falsas y las verdaderas. Estos errores son, esencialmente, dos: a) error de Tipo I y b) error de Tipo II (tabla 4.1). Las dos clases de error tienen diferentes consecuencias y costos, por lo que la probabilidad de cometer ambos errores se intenta mantener tan baja como sea posible.

Tabla 4.1. Probabilidades de error en pruebas de significación estadística (Lipsey, 1990; p. 30).

<u>Conclusión de la prueba estadística</u>	<u>Situación en la población</u>	
	T y C difieren	T y C no difieren
Diferencias significativas (rechazar $H_0$ )	Conclusión correcta Probabilidad = $1-\beta$ (potencia)	Error de Tipo I Probabilidad = $\alpha$
Diferencias no significativas (aceptar $H_0$ )	Error de Tipo II Probabilidad = $\beta$	Conclusión correcta Probabilidad = $1-\alpha$

(T = Grupo de Tratamiento; C = Grupo Control)

El error de Tipo I o nivel de significación ( $\alpha$ ) es el más familiar para la mayoría de investigadores y, por lo común, está fijado en 0.05. Al aplicar una prueba estadística también existe la posibilidad de cometer un error de Tipo II, denotado por  $\beta$ , que es la probabilidad de aceptar la  $H_0$  cuando es falsa. Estos dos errores ( $\alpha$  y  $\beta$ ) están inversamente relacionados. En efecto, al controlar el error de Tipo I, el error de Tipo II queda determinado (Cohen, 1988; Pardo y San Martín, 1994; Rossi, 1990; Stevens, 1986). Así, dada una muestra, al disminuir  $\alpha$  aumenta  $\beta$  y si se desea disminuir ambos errores, la única forma consiste en aumentar el tamaño de muestra, que no siempre es posible. En

consecuencia, el problema que surge al investigador es el de alcanzar un equilibrio entre  $\alpha$  y  $\beta$ . Un análisis de los costes, beneficios y otros factores relevantes asociados a los errores Tipo I y II se halla en Brown (1983), Cascio y Zedeck (1983), Nagel y Neef (1977) y Schneider y Darcy (1984).

La probabilidad de un error Tipo I viene designada de antemano por  $\alpha$ , de modo que el riesgo de rechazar  $H_0$  verdaderas queda controlado. Por contra, la tasa de error Tipo II no se puede determinar hasta que se conozca la  $H_1$ , o sea, hasta que el investigador sepa el efecto concreto existente. Por lo tanto, el experimentador deberá basarse en estimaciones.

El complemento de la probabilidad de un error Tipo II ( $1-\beta$ ) es la potencia de una prueba estadística. Cuando  $\beta$  es pequeño, entonces la potencia es grande y viceversa. En la teoría de Neyman-Pearson, la potencia de una prueba estadística es la probabilidad de resultados significativos; es decir, la aceptación de la  $H_1$  cuando es verdadera (Cohen, 1988; Lipsey, 1990). En otras palabras, la potencia de un estadístico viene dada por su capacidad de rechazar la  $H_0$  correctamente, de modo que está determinada por la probabilidad de cometer errores de Tipo II (tabla 4.1). Cohen (1992) propone por convención una potencia de 0.80 (o sea,  $\beta=0.20$ ). Un valor sustancialmente inferior a 0.80 implicaría un gran riesgo de incurrir en un error Tipo II, y un valor superior exigiría una muestra muy grande, fuera de los recursos del investigador.

Según Strube (1985), el hecho de rechazar la  $H_1$  debido a una potencia baja lleva, en un nivel de meta-análisis, a abandonar prematuramente una área de investigación prometedora. Por ello, dada la influencia que tienen las revisiones de estudios anteriores para una investigación futura, es indispensable que las decisiones concernientes a las hipótesis sean correctas (Sánchez y Ato, 1989).

### 4.3.2. Nivel de significación

El nivel de significación es un estadístico de prueba en virtud del cual se concluye acerca de la existencia el fenómeno, o el riesgo de rechazar erróneamente la  $H_0$ . A no ser que se indique lo contrario, se toma  $\alpha=0.05$  (es parte del legado fisheriano); aunque evidentemente pueden seleccionarse otros valores.

Asumiendo que un fenómeno concreto existe en un determinado grado en la población, cuanto más estricto es el estándar de prueba ( $\alpha$ ), menor es la probabilidad de que la muestra lleve a resultados que se ajusten a dicho estándar (Cohen, 1988; Lipsey, 1990; Pardo y San Martín, 1994; Sedlmeier y Gigerenzer, 1989). Así pues, la práctica de tomar valores  $\alpha$  pequeños -cuanto más pequeños mejor- conduce a una baja potencia. Además, el complemento de la potencia ( $\beta$ ) constituye también un error (el error de Tipo II), por lo que la inferencia estadística puede ser considerada como una ponderación de los dos tipos de error (Cohen, 1962). Por ejemplo, un investigador puede establecer el riesgo de rechazar equivocadamente la  $H_0$  a un nivel de  $\alpha=0.001$  y, con ello, reducir la potencia de la prueba (por consiguiente, aumentar el valor de  $\beta$ ). Por otro lado, es posible incrementar la potencia de la prueba hasta un valor aceptable usando un nivel de significación menos riguroso. No obstante, de esta forma se incrementa la probabilidad de cometer un error Tipo I. Por lo tanto, debe equilibrarse la probabilidad de incurrir en un error de Tipo I con la probabilidad de incidir en un error de Tipo II, puesto que, como ya se ha comentado anteriormente, no puede reducirse la probabilidad de cometer un tipo de error sin que se incremente el otro.

## 5. REFERENCIAS BIBLIOGRÁFICAS

- Adam, J. (1978). Sequential strategies and the separation of age, cohort, and time-of-measurement contributions to developmental data. *Psychological Bulletin*, 85, 1309-1316.
- Anderson, S., Auquier, A., Hauchk, W.W., Oakes, D., Vandealet, W. y Weisberg, H. (1980). *Statistical methods for comparative studies*. New York: John Wiley.
- Arnau, J. (1994a). *Diseños de investigación aplicados*. (Puede obtenerse de J. Arnau, Universidad de Barcelona, Facultad de Psicología, Departamento de Metodología de las Ciencias del Comportamiento, 08035 Barcelona).
- Arnau, J. (1995b). Metodologías cuantitativas en la investigación psicológica. En J. Arnau (Ed.), *Mètodes, dissenys i tècniques en investigació psicològica*. Barcelona: Universitat Oberta de Catalunya.
- Arnau, J. (1995c). Estructura formal del diseño de investigación. En J. Arnau (Ed.), *Diseños longitudinales aplicados a las ciencias sociales y del comportamiento*. México: Limusa.
- Arnau, J. (1995d). Diseños longitudinales de un solo sujeto y una sola variable. En J. Arnau (Ed.), *Diseños longitudinales aplicados a las ciencias sociales y del comportamiento*. México: Limusa.
- Arnau, J. (1995e). Diseños de medidas repetidas de un solo grupo de sujetos. En J. Arnau (Ed.), *Diseños longitudinales aplicados a las ciencias sociales y del comportamiento*. México: Limusa.
- Arnau, J. (1995f). Diseños de medidas repetidas II: diseños de dos grupos y diseños de estructuras más complejas. En J. Arnau (Ed.), *Diseños*
- Arnau, J. (1997). *Diseños de investigación aplicados en esquemas*. Barcelona: Publicacions Universitat de Barcelona.
- Arnau, J. y Gómez, J. (1995). Diseños longitudinales en panel. En J. Arnau (Ed.), *Diseños longitudinales aplicados a las ciencias sociales y del comportamiento*. México: Limusa.
- Ato, M. (1991). *Metodología de la investigación en ciencias del comportamiento I: fundamentos*. Barcelona: PPU-DM.

- Ato, M. (1995b). Conceptos básicos. En M.T. Anguera, J. Arnau, M. Ato, R. Martínez, J. Pascual, G. Vallejo (Eds.), *Métodos de investigación en psicología*. Madrid: Síntesis.
- Ato, M. y Rabadán, R. (1991). *Validez de la investigación psicológica*. Murcia: Torreta, S.A.L.
- Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, 66, 432-437.
- Baltes, P.B., Reese, H.W. y Nesselroade, J.R. (1988). *Introduction to research methods. Life-span developmental psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bentler, P.M. (1978). The interdependence of theory, methodology, and empirical data: Causal modeling as an approach to construct validation. En D.B. Kandel (Ed.), *Longitudinal research on drug use*. Washington: Hemisphere Publishing Corporation.
- Bock, R.D. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.
- Bock, R.D. (1979). Univariate and multivariate analysis of variance of time-structured data. En J.R. Nesselroade y P.B. Baltes (Eds.), *Longitudinal research in the study of behavior and development*. New York: Academic Press.
- Bono, R. y Arnau, J. (1995). Consideraciones generales en torno a los estudios de potencia. *Anales de psicología*, 11, 193-202.
- Box, G.E.P. (1950). Problems in the analysis of growth and wear-curves. *Biometrics*, 6, 363-389.
- Box, G.E.P. y Jenkins, G.M. (1970). *Time-series analysis: Forecasting and control*. San Francisco, CA: Holden-Day.
- Brown, G.W. (1983). Errors, Type I and II. *American Journal of Disorders in Childhood*, 137, 586-591.
- Bryk, A.S. y Weisberg, H.I. (1977). Use of the nonequivalent control group design when subjects are growing. *Psychological Bulletin*, 84, 950-962.
- Campbell, D.T. (1988). *Methodology and epistemology for social science: Selected papers*. Chicago, IL: University of Chicago Press.

- Campbell, D.T. y Boruch, R.F. (1975). Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. En C.A. Bennett y A.A. Lumsdaine (Eds.), *Evaluation and experiment*. New York: Academic Press.
- Campbell, D.T. y Stanley, J.C. (1963). Experimental and quasi-experimental designs for research on teaching. En N.L. Gage (Ed.), *Handbook of research teaching*. Chicago, IL: Rand McNally.
- Campbell, D.T. y Stanley, J.C. (1966). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally. (Traducción española: Amorrortu, Buenos Aires, 1973).
- Cappelleri, J.C., Trochim, W.M.K., Stanley, T.D. y Reichardt, C. (1991). Random measurement error does not bias the treatment effect estimate in the regression-discontinuity design. I. The case of no interaction. *Evaluation Review*, 15, 395-419.
- Cascio, W.F. y Zedeck, S. (1983). Open a new window in rational research planning: Adjust alpha to maximize statistical power. *Personnel Psychology*, 36, 517-526.
- Chase, L.J. y Tucker, R.K. (1976). Statistical power: derivation, development, and data-analytic implications. *The Psychological Record*, 26, 473-486.
- Cochran, W.G. (1968). Error of measurement in statistics. *Technometrics*, 10, 637-666.
- Cohen, D.K. (1975). The value of social experiments. En A.M. Rivlin y P.M. Timpane (Eds.), *Planned variation in education: Should we give up or try harder?*. Washington, DC: Brookings Institution.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2ª ed.). Hillsdale, NJ: Erlbaum.
- Cook, T.D. y Campbell, D.T. (1976). The design and conduct of quasi-experiments and true experiments in field settings. En M. Dunnette (Ed.), *Handbook of industrial and organizational psychology*. Skokie, IL: Rand McNally.
- Cook, T.D. y Campbell, D.T. (1979). *Quasi-experimentation. Design and analysis issues for field settings*. Chicago, IL: Rand McNally.
- Cook, T.D. y Campbell, D.T. (1986). The causal assumptions of quasi-experimental practice. *Synthese*, 68, 141-180.

- Cook, T.D., Campbell, D.T. y Peracchio, L. (1990). Quasi experimentation. En M.D. Dunnette y L.M. Hough (Eds.), *Handbook of industrial and organizational psychology*. Palo Alto, CA: Consulting Psychologist Press.
- Costa, P.T. y McCrae, R.R. (1982). An approach to the attribution of aging, period, and cohort effects. *Psychological Bulletin*, 91, 238-250.
- Cox, D.R. (1957). The use of a concomitant variable in selecting an experimental design. *Biometrika*, 44, 150-158.
- Donaldson, G. y Horn, J.L. (1992). Age, cohort, and time developmental muddles: Easy in practice, hard in theory. *Experimental Aging Research*, 18, 213-222
- Dwyer, J.H. (1983). *Statistical models for the social and behavioral sciences*. New York: Oxford University Press.
- Edwards, A.L. (1984). *Experimental design in psychology research* (5<sup>a</sup> ed.). New York: Harper and Row.
- Feldt, L.S. (1958). A comparison of the precision of three experimental designs employing a concomitant variable. *Psychometrika*, 23, 335-353.
- Fisher, R.A (1935). *The design of experiments*. London: Oliver and Boyd.
- Fisher, R.A (1950). *Contributions to mathematical statistics*. London: Chapman and Hall.
- Fisher, R.A (1955). Statistical methods and scientific induction. *Journal of the Royal Statistical Society, Series B*, 17, 245-251.
- Games, P.A. (1990). Alternative analyses of repeated-measure designs by ANOVA and MANOVA. En A. von Eye (Ed.), *Statistical methods in longitudinal research*, Vol 1. San Diego, CA: Academic Press.
- Gigerenzer, G. y Murray, D.J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.
- Gilbert, J.P., Light, R.J. y Mosteller, F. (1975). Assessing social innovation: An empirical base for social policy. En C.A. Bennett y A.A. Lumsdaine (Eds.), *Evaluation and experimental: Some critical issues in assessing social programs*. New York: Academic Press.
- Glass, G., Willson, V.L. y Gottman, J.M. (1975). *Design and analysis of time-series experiments*. Boulder, CO: Colorado Associated University Press.
- Glenn, N.D. (1977). *Cohort analysis. Series: Quantitative applications in the social sciences*. Newbury Park, CA: Sage.



- Goldstein, H. (1979). *The design and analysis of longitudinal studies*. London: Academic Press.
- Gottman, J.M. (1981). *Time-series analysis*. Cambridge: Cambridge University Press.
- Greenhouse, S.W. y Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95-112.
- Hagenaars, J.A. (1990). *Categorical longitudinal data: Log-linear panel, trend, and cohort analysis*. Newbury Park, CA: Sage.
- Hedrick, T.E., Bickman, L. y Rog, D.J. (1993). *Applied research design. A practical guide*. Newbury Park, CA: Sage.
- Horn, J.L. y Donaldson, G. (1977). Faith is not enough: A response to the Baltes-Schaie claim that intelligence does not wane. *American Psychologist*, 32, 369-373.
- Hsiao, C. (1986). *Analysis of panel data*. Cambridge: Cambridge University Press.
- Huitema, B.E. (1980). *The analysis of covariance and alternatives*. New York: John Wiley.
- Huynh, H. y Feldt, L.S. (1970). Conditions under which mean square ratios in repeated measurements designs have exact *F* distributions. *Journal of the American Statistical Association*, 1, 1582-1589.
- Judd, C.M. y Kenny, D.A. (1981). *Estimating the effects of social interventions*. Cambridge: Cambridge University Press.
- Judd, C.M. y McClelland, G.H. (1989). *Data analysis: A model-comparison approach*. San Diego, CA: Harcourt Brace Jovanovich.
- Judd, C.M., McClelland, G.H. y Culhane, S.E. (1995). Data analysis: continuing issues in the everyday analysis of psychological data. *Annual Review of Psychology*, 46, 433-465.
- Kenny, D.A. (1975a). A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design. *Psychological Bulletin*, 82, 345-362.
- Kenny, D.A. (1979). *Correlation and causality*. New York: John Wiley.
- Kirk, R.E. (1982). *Experimental designs: Procedures for the behavioral sciences* (2<sup>a</sup> ed.). Monterey, CA: Brooks/Cole.
- Lipsey, M.W. (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.

- Lord, F.M. (1960). Large-sample covariance analysis when the control variable is fallible. *Journal of the American Statistical Association*, 55, 307-321.
- Lord, F.M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, 68, 304-305.
- Mark, M.M. (1986). Validity typologies and the logic and practice of quasi-experimentation. En W.M.K. Trochim (Ed.), *Advances in quasi-experimental design and analysis*. San Francisco, CA: Jossey-Bass.
- Markus, G.B. (1985). Dynamic modeling of cohort change: The case of political partisanship. En W.M. Mason y S.E. Fienberg (Eds.), *Cohort analysis in social research*. New York: Springer Verlag.
- Mauchley, J.W. (1940). Significance test of sphericity of a normal  $n$ -variate distribution. *Annals of Mathematical Statistics*, 11, 204-209.
- Maxwell, S.E. y Delaney, H.D. (1990). *Designing experiments and analyzing data. A model comparison perspective*. Belmont, CA: Wadsworth.
- McCleary, R. y Hay, A. (1980). *Applied time series analysis for the social sciences*. Beverly Hills, CA: Sage.
- Menard, S. (1991). *Longitudinal research*. Beverly Hills, CA: Sage.
- Mohr, L.B. (1982). On rescuing the nonequivalent-control-group design. The random-comparison-group approach. *Sociological Methods and Research*, 11, 53-80.
- Morrison, D.F. (1967). *Multivariate statistical methods*. New York: McGraw-Hill.
- Nagel, S.S. y Neef, M. (1977). Determining an optimum level of statistical significance. En M. Guttentag y S. Saar (Eds.), *Evaluation studies review annual*, Vol 2. Beverly Hills, CA: Sage.
- Neter, J. Wasserman, W. y Kutner, M.H. (1985). *Applied linear statistical models: Regression, analysis of variance and experimental designs* (2<sup>a</sup> ed.). Homewood, IL: Richard D. Irwin.
- Neyman, J. y Pearson, E.S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, 20A, 175-240.
- Neyman, J. y Pearson, E.S. (1933). On the problem of the most efficient tests of statistical hypothesis. *Transactions of the Royal Society of London Series A*, 231, 289-337.
- Neyman, J. y Pearson, E.S. (1936). Contributions to the theory of testing statistical hypotheses. *Statistical Research Memoirs*, 1, 1-37.

- Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioral sciences*. New York: Wiley.
- Pardo, A. y San Martín, R. (1994). *Análisis de datos en psicología II*. Madrid: Pirámide.
- Pascual, J. (1995). Diseño de medidas repetidas. En M.T. Anguera, J. Arnau, M. Ato, R. Martínez, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología*. Madrid: Síntesis.
- Pedhazur, E.J. y Schmelkin, L.P. (1991). *Measurement, design, and analysis. An integrated approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Peters, C.C. y Van Voorhis, W.R. (1940). *Statistical procedures and the mathematical bases*. New York: McGraw-Hill.
- Plewis, I. (1985). *Analysing change*. New York: John Wiley.
- Porter, A.C. (1967). The effects of using fallible variables in the analysis of covariance, doctoral dissertation, University of Wisconsin, Madison.
- Porter, A.C. y Chibucos, T.R. (1974). Selecting analysis strategies. En G.D. Borich (Ed.), *Evaluating educational programs and products*. Englewood Cliffs, NJ: Educational Technology Publications.
- Reichardt, C.S. (1979). The statistical analysis of data from nonequivalent group designs. En T.D. Cook y D.T. Campbell, *Quasi-experimentation. Design and analysis issues for field settings*. Chicago, IL: Rand McNally.
- Reichardt, C.S., Trochim, W.M.K., Cappelleri, J.C. (1995). Reports of the death of regression-discontinuity analysis are greatly exaggerated. *Evaluation Review*, 19, 39-63.
- Reynolds, A.J. y Temple, J.A. (1995). Quasi-experimental estimates of the effects of a preschool intervention. Psychometric and econometric comparisons. *Evaluation Review*, 19, 347-373.
- Riba, M.D. (1987). *Modelo lineal de análisis de la variancia*. Bellaterra: Universidad Autónoma de Barcelona.
- Riba, M.D. (1990). *Modelo lineal de análisis de la variancia*. Barcelona: Herder.
- Riecken, H.W. y Boruch, R.F. (Eds.) (1974). *Social experimentation: A method for planning and evaluating social intervention*. New York: Academic Press.
- Rogosa, D. (1979). Causal models in longitudinal research: Rationale, formulation, and interpretation. En J.R. Nesselroade y P.B. Baltes (Eds.), *Longitudinal research in the study of behavior and development*. New York: Academic Press.

- Rogosa, D. (1980a). Comparing non-parallel regression lines. *Psychological Bulletin*, 88, 307-321.
- Rogosa, D., Brandt, D. y Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, 92, 726-748.
- Rossi, J.S. (1990). Statistical power of psychological research: What have we gained in 20 years?. *Journal of Consulting and Clinical Psychology*, 58, 646-656.
- Rossi, P.H. y Freeman, H.E. (1989). *Evaluation. A systematic approach* (3ª ed.). Beverly Hills, CA: Sage. (Traducción española: Trillas, México, 1989).
- Rouanet, H. y Lepine, D. (1970). Comparison between treatments in a repeated measures design: ANOVA and multivariate methods. *British Journal of Mathematical and Statistical Psychology*, 23, 147-163.
- Rubin, D.B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*, 2, 1-26.
- Rutter, M. (1988). Longitudinal data in the study of causal processes: Some uses and some pitfalls. En M. Rutter (Ed.), *Studies of psychological risk: The power of longitudinal data*. Cambridge: Cambridge University Press.
- Ryder, N.B. (1965). The cohort as a concept in the study of social change. *American Review of Sociology*, 30, 843-861.
- Sánchez, J. y Ato, M. (1989). Meta-análisis: una alternativa metodológica a las revisiones tradicionales de la investigación. En J. Mayor y J.L. Pinillos (Eds.), *Tratado de psicología general I. Historia, teoría y método*. Madrid: Alhambra.
- Schaie, K.W. (1965). A general model for the study of developmental problems. *Psychological Bulletin*, 64, 92-107.
- Schaie, K.W. (1970). A reinterpretation of age-related changes in cognitive structure and functioning. En L.R. Goulet y P.B. Baltes (Eds.), *Life-span developmental psychology: Research and theory*. New York: Academic Press.
- Schaie, K.W. (1972). Limitations on the generalizability of growth curves of intelligence: A reanalysis of some data from the Harvard Growth Study. *Human Development*, 15, 141-152.
- Schaie, K.W. (1974). Translations in gerontology -from lab to life: Intellectual functioning. *American Psychologist*, 29, 802-807.
- Schaie, K.W. (1992). The impact of methodological changes in gerontology. *International Journal of Aging and Human Development*, 35, 19-29.

- Scheirs, J.G.M. (1992). A priori and a posteriori test on repeated measurements. *Educational Psychology, 12*, 63-72.
- Schneider, A.L. y Darcy, R.E. (1984). Policy implications of using significance tests in evaluation research. *Evaluation Review, 8*, 573-582.
- Seber, G.A.F. (1977). *Linear regression analysis*. New York: John Wiley.
- Seber, G.A.F. (1980). *The linear hypothesis: A general theory* (2ª ed.). London: Griffin.
- Sedlmeier, P. y Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies?. *Psychological Bulletin, 105*, 309-316.
- Stanley, T.D. (1991). "Regression-discontinuity design" by any other name might be less problematic. *Evaluation Review, 15*, 605-624.
- Stanley, T.D. y Robinson, A. (1990). Sifting statistical significance from the artifact of regression-discontinuity design. *Evaluation Review, 14*, 166-181.
- Stevens, J. (1986). *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Stoolmiller, M. (1995). Using latent growth curve models to study developmental processes. En J.M. Gottman (Ed.), *The analysis of change*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Strube, M.J. (1985). Power analysis for combining significance levels. *Psychological Bulletin, 98*, 595-599.
- Suchman, E.A. (1967). *Evaluative research*. New York: Russell Sage Foundation.
- Trochim, W.M.K. (1984). *Research design for program evaluation: The regression-discontinuity approach*. Beverly Hills, CA: Sage.
- Trochim, W.M.K., Cappelleri, J.C. y Reichardt, C.S. (1991). Random measurement error does not bias the treatment effect estimate in the regression-discontinuity design. II. When an interaction effect is present. *Evaluation Review, 15*, 571-604.
- Tukey, J.W. (1991). The philosophy of multiple comparisons. *Statistical Science, 6*, 100-116.
- Vallejo, G. (1996). *Diseño de series temporales interrumpidas*. Barcelona: Ariel
- Van Dalen, D.B. y Meyer, W.J. (1971). *Manual de técnica de la investigación educacional*. Buenos Aires: Paidós.
- Visser, R.A. (1985). *Analysis of longitudinal data in behavioural and social research*. Leiden: DSWO Press.

- Wall, W.D. y Williams, H.L. (1970). *Longitudinal studies and the social sciences*. London: Heinemann.
- Weisberg, H.I. (1979). Statistical adjustments and uncontrolled studies. *Psychological Bulletin*, 86, 1149-1164.
- Weiss, C.H. (1972b). Evaluating educational and social action programs: A "treeful of owls". En C.H. Weiss (Ed.), *Evaluatin action programs: Readings in social action and research*. Boston, MA: Allyn and Bacon.
- Williams, J.D. (1991). Age, cohort and period in life-span research: A three-way analysis with logically missing cells. *Multivariate Behavioral Research*, 26, 631-654.
- Winer, B.J. (1971). *Statistical principles in experimental design* (2<sup>a</sup> ed.). New York: McGraw-Hill.