

DISEÑOS DE INVESTIGACIÓN EN PSICOLOGÍA

Prof. M. Isabel Núñez Peña

**Departamento de Metodología de las Ciencias del Comportamiento
Facultad de Psicología
Universidad de Barcelona**

ÍNDICE

2. Diseños de investigación en Psicología	1
2.1. La investigación científica	1
2.2. Planificación de la investigación científica	6
2.2.1. Nivel teórico-conceptual.....	6
2.2.2. Nivel técnico-metodológico.....	8
2.2.3. Nivel analítico-estadístico.....	11
2.3. Tipos de diseños	13
2.3.1. Diseños experimentales clásicos	14
2.3.1.1. Definición y características	14
2.3.1.2. Clasificación de los diseños experimentales clásicos	17
2.3.1.3. Análisis de datos para este tipo de diseños	22
2.3.2. Diseños experimentales de N=1	34
2.3.2.1. Definición y características	34
2.3.2.2. Clasificación de los diseños de N=1.....	39
2.3.2.3. Análisis de datos para este tipo de diseños	47
2.3.3. Diseños cuasi-experimentales.....	52
2.3.3.1. Definición y características	52
2.3.3.2. Clasificación de los diseños cuasi-experimentales.....	57
2.3.3.3. Análisis de datos para este tipo de diseños	60
2.3.4. Diseños no-experimentales	63
2.3.4.1. Definición y características	63
2.3.4.2. Clasificación de los diseños no-experimentales	65
2.3.4.3. Análisis de datos para este tipo de diseños	69
Referencias bibliográficas.....	71

2. DISEÑOS DE INVESTIGACIÓN EN PSICOLOGÍA

2.1. LA INVESTIGACIÓN CIENTÍFICA

Dado que la materia que nos ocupa es el diseño de investigaciones científicas en Psicología, consideramos que una buena manera que comenzar es haciendo una reflexión inicial acerca de qué es la Ciencia, qué es la investigación científica y cuáles son los paradigmas de investigación científica en Psicología.

Comenzaremos con el concepto de ciencia. El término ciencia se define como *conjunto organizado de conocimientos que versa sobre un objeto determinado, tiene un método propio y se funda en relaciones objetivas comprobables* (Seco, 1999). Atendiendo a esta definición, vemos que la Ciencia se caracteriza por su objeto y por su método. Conviene, pues, conocer más detalles de uno y de otro.

El objetivo de la ciencia es explicar lo que acontece en el mundo que nos rodea ofreciendo explicaciones sistemáticas de aquellas realidades que estudia (Pizarro, 1986). En otras palabras: se trata de establecer reglas, más o menos generales, sobre el funcionamiento de la naturaleza. En el caso de la Psicología, lo que se pretende es establecer leyes generales y ciertas que nos permitan alcanzar, en la medida de lo posible, una mayor y mejor comprensión del comportamiento del ser humano.

En el párrafo anterior decimos que el objetivo básico de la ciencia es elaborar explicaciones de los fenómenos que acontecen. Esto quiere decir que la ciencia da respuesta a preguntas acerca de qué es lo que está ocurriendo y por qué está ocurriendo. Las respuestas a estas preguntas son las explicaciones del porqué de los

fenómenos, o, lo que es lo mismo, sus causas. ¿Para que nos sirve conocer las causas? Si comprendemos por qué ocurren los fenómenos podremos llegar a predecirlos, es decir, anticiparnos a ellos. En consecuencia, se podrá incluso intervenir en el mundo para mejorar el futuro. Ya a principios de siglo XVII, Francis Bacon propuso que la finalidad de la ciencia es la mejora de la suerte del hombre en la tierra, y, según él, esa finalidad se lograría recogiendo hechos a través de la observación organizada y derivando de ellos teorías.

Para alcanzar su objetivo, la Ciencia procede de una manera especial que es utilizando el método científico. Si nos remontamos al origen etimológico de la palabra método comprobaremos que viene de la fusión de las palabras griegas *metá* y *odós*, que significan, respectivamente, hacia y camino. Así, literalmente, el método es el camino que nos lleva hacia un fin. En el caso de la ciencia, se trata de un camino que se caracteriza por la sistematicidad, la objetividad y la rigurosidad. Este método, seguido en el razonamiento científico, se conoce como método hipotético-deductivo.

Una vez se ha aclarado qué es la ciencia, cuál es su objeto y cuál es su método vamos a proceder con la definición de la investigación científica. Se trata de un proceso de indagación que persigue la elaboración de explicaciones sistemáticas, objetivas y rigurosas de los fenómenos, y que para ello recurre al método científico. Señala Kirk (1995) que no todos los fenómenos pueden abordarse desde un punto de vista científico, puesto que los objetos de investigación científica deben ser comprobables por la observación o producidos por medio de un experimento. Con esto quiere decir que hay cuestiones que no son susceptibles de ser investigadas científicamente, como, por ejemplo, la existencia de vida en otras galaxias. En este caso la razón es muy clara: por el momento no disponemos de tecnologías para observar otras galaxias distintas a la nuestra.

Hay otra característica que deben presentar los fenómenos para poderlos investigar: sólo podrán ser objeto de la investigación científica aquellos fenómenos para los cuales pueda repetirse su observación o puedan volver a producirse. Esto quiere decir que un científico debe asegurarse de que sus resultados puedan ser reproducidos por otros investigadores.

Para acabar, vamos a presentar los paradigmas metodológicos o tradiciones de la investigación científica en Psicología. Cronbach (1957) en el discurso que realizó en la 65ª convención anual de la Asociación Americana de Psicología en Nueva York describió dos métodos de investigación propios de la Psicología: el método

experimental y el método correlacional. Por su parte, Arnau (1995a) nos dice que en la ciencia psicológica están presentes dos paradigmas: el paradigma experimental y el paradigma asociativo (puede verse un esquema de éstos en la Figura 2.1). Según Arnau (1995a) cada paradigma se caracteriza por: (1) la formulación de una clase específica de hipótesis, (2) el grado de intervención del investigador en la situación estudiada, y (3) los procedimientos de recogida de datos.

Por lo que se refiere al tipo de hipótesis, en el paradigma experimental se formulan hipótesis causales, mientras que, en el paradigma asociativo se formulan hipótesis de covariación. Una hipótesis causal puede definirse como un enunciado que establece una relación inequívoca de causa-efecto. En términos de variables, esta hipótesis establece una relación entre la variable independiente y la variable dependiente, y presenta la forma lógica *Si A entonces B*. Una hipótesis de covariación, por el contrario, es un enunciado en el que se establece una relación asociativa (o no-causal) entre fenómenos. Esto quiere decir que este tipo de hipótesis nos informa de que la variable A está asociada o relacionada con la variable B. Obviamente, una hipótesis causal implicará relación entre las variables, sin embargo, una hipótesis de covariación no necesariamente implicará causalidad.

Según Kenny (1979) una relación será causal si se dan tres condiciones:

1. Precedencia temporal de la causa.
2. Relación funcional entre las variables o covariación causa-efecto, de manera que cuando se da la presunta causa aparece el presunto efecto y cuando no se da la presunta causa no aparece el presunto efecto.
3. Ausencia de espuriedad, que implica el control previo de las variables extrañas. Es decir, no debe haber otras explicaciones plausibles del cambio.

El segundo elemento diferenciador de paradigmas es el grado de intervención del investigador. En el paradigma experimental el grado de intervención es máximo. Esto se debe a que el objetivo de este tipo de investigaciones es establecer relaciones causales inequívocas entre las variables y para ello el fenómeno se ha de observar en un contexto muy controlado. El experimentador deberá actuar de la siguiente manera: (1) provocará el fenómeno que quiere estudiar y (2) lo aislará para que ningún elemento extraño afecte a la relación causal que está estudiando.

Dicho en términos más técnicos: el investigador manipulará la variable independiente y controlará el efecto de las variables de confundido (también denominadas, variables extrañas). Manipular una la variable independiente quiere decir que se va a incidir directamente sobre el fenómeno objeto de la investigación con la intencionalidad de producir un cambio sobre el mismo. Controlar el efecto de variables de confundido quiere decir que debemos garantizar que los cambios observados en el fenómeno se deban a la acción directa del experimentador y no a la acción de otras variables. Este alto grado de intervención del investigador sobre el fenómeno es lo que ha originado la principal crítica que recibe el paradigma experimental; a saber: su artificialidad. Los fenómenos en la realidad no se producen en situaciones tan controladas, por lo que se puede cuestionar que las conclusiones elaboradas en este paradigma sean válidas en contextos naturales. Sin embargo, algunos autores como Pereda (1987) nos recuerdan que siempre será preferible saber algo, aunque no se pueda generalizar, que generalizar algo que no se sabe.

El paradigma asociativo se caracteriza porque el grado de intervención del investigador es mínimo. Esto quiere decir que el investigador ni provoca el fenómeno que quiere estudiar ni lo aísla, sino que se limita a estudiarlo mientras éste sigue su curso natural. En términos metodológicos esto se traduce en la ausencia de manipulación de la variable independiente y en un mínimo o nulo control de las fuentes de variación extrañas. En este sentido las investigaciones que se realizan dentro del paradigma asociativo sortean la principal crítica que recibe el paradigma experimental: su artificialidad. No obstante, presentan un inconveniente que es la imposibilidad de establecer relaciones causales.

El tercer elemento que diferencia entre los dos paradigmas es el procedimiento de recogida de datos. En otras palabras, los paradigmas se diferencian en cuanto al tipo de diseños que utilizan. Arnau (1995a) señala que los diseños experimentales y los cuasi-experimentales se utilizan dentro del paradigma experimental, mientras que los diseños observacionales y los diseños de encuesta se utilizan dentro del paradigma asociativo. Aunque ahora no pretendemos extendernos en exceso en la descripción de los tipos diseños —tema que se abordaremos más adelante—, creemos que es conveniente dar algunas pinceladas respecto a éstos.

De lo anterior se desprende que tanto los diseños experimentales como los diseños cuasi-experimentales se utilizan en investigaciones que tienen como objetivo el establecimiento de relaciones causales inequívocas. En consecuencia, en

ambos tipos de diseños se recurre a la manipulación de una o varias variables independientes y al control de las variables extrañas. ¿Qué es, entonces, lo que los diferencia? El grado de control de las variables extrañas. Estas diferencias en cuanto al grado de control se producen porque los diseños cuasi-experimentales utilizan grupos no equivalentes —también denominados, grupos intactos o naturales—, mientras que los diseños experimentales utilizan grupos equivalentes —es decir, grupos que se forman al azar—.

Por lo que se refiere a los diseños observacionales y al diseño de encuestas, éstos se incluyen en el paradigma asociativo (Arnau, 1995a). Se trata de metodologías mediante las que el investigador recaba sus datos sin intentar inducir cambios o introducir tratamientos. La diferencia fundamental entre una y otra radica en que la primera se ocupa de investigar fenómenos directamente observables —en general, la conducta de los individuos—, mientras que la segunda se interesa por el estudio de procesos que no pueden abordarse mediante la observación directa —por ejemplo, opiniones o actitudes—. Esta diferencia repercute en el tipo de técnicas que se utilizan en cada caso para recoger información. De una parte, la metodología observacional se sirve de técnicas sistemáticas de observación de la conducta. De otra parte, la metodología de encuestas emplea cuestionarios, a través de los cuales se obtiene información sobre opiniones, actitudes o determinados constructos psicológicos.

Tabla 2.1. Paradigmas de investigación en Psicología (Arnau, 1995a)

Paradigma experimental	Paradigma asociativo
Hipótesis causal	Hipótesis de covariación
Manipulación de la variable independiente	No manipulación de la variable independiente
Elevado control de fuentes de variación extrañas	Mínimo o nulo control de fuentes de variación extrañas
Diseños experimentales y cuasi-experimentales	Diseños observacionales y diseños de encuesta

2.2. PLANIFICACIÓN DE LA INVESTIGACIÓN CIENTÍFICA

Todo investigador seguirá una serie de etapas para realizar su investigación: 1º) planteará el problema de la investigación; 2º) formulará la hipótesis o conjetura; 3º) seleccionará la estrategia para solventar el problema; 4º) recogerá los datos; 5º) analizará los datos; 6º) interpretará los resultados estadísticos; y 7º) elaborará las conclusiones. Dada la naturaleza cíclica de la ciencia, estas conclusiones le llevarán a él o a otros investigadores a plantear nuevos problemas de investigación. Como señalan Box, Hunter y Hunter (1978), el desarrollo de la ciencia es siempre progresivo a través de un proceso cíclico de aprendizaje.

Las etapas anteriores se reparten en tres niveles de actuación que están presentes en cualquier proceso de la investigación científica (Arnau, 1989, 1990): el nivel teórico-conceptual, el nivel técnico-metodológico, y el nivel analítico-estadístico. Dentro del nivel teórico-conceptual se incluyen el planteamiento del problema, la formulación de la hipótesis y la elaboración de conclusiones. Dentro del nivel técnico-metodológico se incluyen las fases correspondientes a la selección de la estrategia para solventar el problema y la recogida de datos. Por último, dentro del nivel analítico-estadístico se incluyen las fases de análisis e interpretación de los resultados estadísticos.

A continuación vamos a profundizar en cada uno de estos niveles y en las etapas de la investigación científica que incluyen.

2.2.1. Nivel teórico-conceptual

Los científicos invierten una buena parte de sus vidas intentando dar respuesta a problemas de investigación, por lo que el punto de partida de sus investigaciones — en definitiva, el origen de la investigación— siempre será una pregunta. En palabras de Pedhazur y Pedhazur Schmelkin (1991) *un problema es un enunciado, usualmente interrogativo, acerca de la relación entre dos o más variables* (p. 187). ¿Es eficaz este nuevo fármaco para el tratamiento del trastorno bipolar? ¿Existen diferencias entre hombres y mujeres en cuanto a su aptitud numérica? ¿Podemos situar el procesamiento de las emociones en alguna región concreta del cerebro? ¿Inciden los prejuicios en los veredictos de los jurados populares? Estos ejemplos recogen problemas susceptibles de ser estudiados científicamente. En este

momento conviene recordar que la ciencia sólo se ocupa de aquellos fenómenos que puedan observarse o puedan producirse por medio de un experimento. En consecuencia, una de las características del problema científico es que éste deberá poder solventarse; es decir, deberemos poder encontrar una solución al mismo en el mundo empírico. Con esto quedan descartados otros problemas o preguntas que puedan formularse pero que no se podrán plantear como problema de investigación científica.

El problema de investigación, además de ser resoluble, deberá presentar otras características. Primero, deberá tener relevancia teórica o práctica (puede encontrarse una discusión de este tema en Pedhazur y Pedhazur Schmelkin, 1991); segundo, deberemos disponer de los recursos necesarios para solventarlo; y, tercero, se deberá solucionar respetando los derechos de los participantes en la investigación (véase en Diener y Crandall, 1978, una amplia exposición sobre el tratamiento ético en la investigación).

Una vez que el científico haya planteado su problema de investigación, el siguiente paso que habrá de dar es formular su hipótesis. La hipótesis es la respuesta al problema de investigación, es decir, recoge la predicción del investigador —los resultados que espera obtener—. Si hemos definido el problema de investigación como un enunciado, usualmente interrogativo, acerca de la relación entre dos o más variables, la hipótesis será el enunciado conjetural en el que se concreta el tipo de relación que esperamos encontrar entre éstas. Así pues, los enunciados de las hipótesis contienen dos o más variables —que son medibles o pueden serlo— y la relación entre éstas. Esta relación se entiende en un sentido amplio, puesto que en las hipótesis se formularán tanto relaciones de causa-efecto como relaciones de covariación. En este sentido, Keppel (1991) cuando se refiere a las hipótesis de investigación nos dice:

These hypotheses may represent deductions or derivations from a more or less formal theoretical explanation of the behavioural phenomenon you wish to study, or they may represent simply hunches or speculations that you tentatively hold concerning this phenomenon (p. 4).

A continuación presentaremos las principales características de las hipótesis (Kerlinger y Lee, 2002):

1. Dirigen la investigación.
2. Se deducen de las teorías o de un conjunto de investigaciones previas.
3. Se pueden someter a prueba: es decir, se puede demostrar que son probablemente verdaderas o probablemente falsas.
4. Son predicciones del tipo *Si A entonces B*, donde se establece la relación entre las variables A y B.
5. Son independientes de los valores y opiniones de una persona.

2.2.2. Nivel técnico-metodológico

Una vez formulada la hipótesis, el investigador se adentrará en el nivel técnico-metodológico. En este nivel lo que deberá hacer es seleccionar la estrategia metodológica más adecuada para contrastar su hipótesis. Es decir: deberá decidir cuál es la mejor manera de proceder para recoger datos que le permitan solucionar el problema de la investigación. Tendrá que responder a las siguientes preguntas: cómo se seleccionará la muestra, cómo se formarán los grupos, cómo se manipulará la variable independiente, cómo se medirá la variable dependiente, cómo se controlarán las variables de confundido, etc. En esta fase de la investigación, es muy importante que el investigador tome las decisiones adecuadas, porque, en caso contrario, puede poner en peligro la validez de su investigación. Las decisiones que tome repercutirán en la validez interna, en la validez externa y en la validez de constructo de su investigación (una amplia descripción de las amenazas a la validez de las investigaciones se encuentra en Cook y Campbell, 1979).

Hay básicamente tres maneras de dar respuesta a los problemas de investigación, o, dicho de otro modo, hay tres estrategias que podemos adoptar para recoger los datos:

1. Podemos manipular algún aspecto del ambiente en una situación totalmente controlada y esperar a ver qué ocurre. En este caso, utilizaremos un diseño experimental —también denominado, experimento verdadero (Cook y Campbell, 1979; Dwyer, 1983)—.
2. Podemos manipular algún aspecto del ambiente en una situación no totalmente controlada y esperar a ver qué ocurre. En este caso, utilizaremos un diseño cuasi-experimental.

3. Podemos observar el fenómeno que nos interesa de forma natural sin interferir con el mismo. En este caso, deberemos recurrir a un diseño no experimental.

Todos los diseños anteriores comparten las siguientes características (Field y Hole, 2003):

1. Empirismo: los tres tipos de diseños intentan obtener evidencias a través de la observación y la medida.
2. Medida: todos los métodos intentan medir de la forma más precisa posible aquello que están estudiando.
3. Replicabilidad: todos buscan asegurarse de que sus resultados puedan ser reproducidos por otros investigadores.
4. Objetividad: todos intentan dar respuesta al problema de investigación de forma objetiva, es decir, independientemente de la interpretación del investigador.

Vamos a ver ahora, a grandes rasgos, cuáles son los principales elementos que diferencian estos tres tipos de diseños. Estos elementos son la existencia de manipulación de la variable independiente, el uso de la aleatorización en la formación de los grupos y el control de las variables de posible confundido. Un experimento verdadero se caracteriza porque en él se manipula la variable independiente, se recurre a la aleatorización para formar grupos equivalentes y se ejerce un control máximo de las variables de posible confundido. Un cuasi-experimento, por el contrario, presenta las mismas características que el experimento verdadero, pero, a diferencia de éste, trabaja con grupos no equivalentes —esto es: grupos naturales o intactos—. Por último, los diseños no experimentales son aquellos en los que no se manipula la variable independiente, hay un control escaso o nulo de las variables de posible confundido y no se recurre a la aleatorización para formar los grupos. En la Tabla 2.2 se presenta un esquema con las principales diferencias entre un experimento verdadero, un cuasi-experimento y un no experimento. Este esquema amplía el esquema de la Tabla 2.1, donde se muestran las características diferenciales entre el paradigma experimental y el paradigma asociativo.

Tabla 2.2. Tipos de diseños

	Experimento verdadero	Cuasi-experimento	No experimento
Manipulación de la variable independiente	SÍ	SÍ	NO
Aleatorización en la formación de grupos	SÍ	NO	NO
Control de variables de confundido	SÍ	SÍ	NO

A la hora de decidir cómo procederá para recoger los datos, el investigador también deberá atender a una dimensión temporal. Queremos decir con esto que le puede interesar estudiar el fenómeno en un momento temporal concreto —por lo que deberá recurrir a un diseño transversal o estático— o, por el contrario, puede interesarle estudiar la evolución del fenómeno a lo largo del tiempo —por lo que deberá recurrir a un diseño longitudinal—. Los diseños longitudinales se utilizan cuando se estudian problemas de investigación que involucran tendencias, cambios o desarrollos a través del tiempo.

Tal como hemos indicado anteriormente, éste es el momento en el que el investigador deberá enfrentarse al tema de la validez de su investigación. La validez significa esencialmente *verdad* o *corrección* (Maxwell y Delaney, 1990). En concreto, el investigador se deberá preocupar de la validez interna, la validez externa y la validez de constructo. Cook y Campbell (1979) nos dicen que la validez interna está relacionada con la validez causal. Esto quiere decir que una investigación tendrá validez interna si se demuestra que los cambios observados en la variable dependiente se deben exclusivamente a la manipulación de la variable independiente; es decir, debe quedar totalmente descartada cualquier explicación alternativa para ese fenómeno. Por su parte, la validez externa en una investigación garantiza que los resultados obtenidos pueden generalizarse a otros sujetos, a otras situaciones y a otros momentos temporales. Por último, la validez de constructo se ocupa de la cuestión de si realmente estamos midiendo el constructo teórico que pretendemos medir. El término constructo se entiende como sinónimo de concepto o de construcción teórica, por lo que se trata de fenómenos que no podemos

observar directamente. Así, una investigación con validez de constructo nos garantiza que las variables observadas son buenos indicadores del constructo o fenómeno no observable que pretendemos medir.

2.2.3. Nivel analítico-estadístico

En el nivel analítico-estadístico el investigador cubrirá las etapas del análisis de los datos y la interpretación de los resultados estadísticos. En este momento, deberá organizar los datos en una tabla o matriz y deberá ser capaz de recurrir a las técnicas estadísticas más adecuadas para obtener información de ésta. Se trata, en definitiva, de ser capaz de extraer conocimiento de esos datos. Hemos afirmado en otro lugar que *todo buen investigador debe tener sólidos y precisos conocimientos sobre estadística* (Solanas, Salafranca, Fauquet y Núñez, 2005, p. 1). Esto es así porque todas las decisiones que el investigador haya tomado en el nivel técnico-metodológico —por ejemplo, cómo se ha seleccionado la muestra, qué tipo de diseño se ha utilizado, el número de variables independientes, el número de variables dependientes, la escala de medida de las variables, etcétera— delimitarán las opciones del análisis estadístico del que se habrá de valer para resolver el problema de su investigación. De hecho, un científico que conozca bien la estadística la utilizará para tomar decisiones no sólo en el nivel analítico-estadístico, sino también en las etapas previas.

Cuando el investigador se enfrenta a la etapa del análisis de datos pondrá en juego todos sus conocimientos de estadística. Las técnicas de análisis a las que podrá recurrir dependerán básicamente de cuatro factores: (1) el objetivo de su análisis, (2) el número de variables que pretenda analizar, (3) la escala en la que haya medido sus variables y (4) la naturaleza de sus datos en cuanto a la posibilidad de existencia de covariación en las medidas.

En primer lugar, dependiendo del objetivo del análisis, se recurrirá a las técnicas de estadística inferencial o a las técnicas de estadística descriptiva. Se utilizarán las primeras si el objetivo es elaborar conclusiones para toda la población a partir de los datos de una muestra. Se trata, en definitiva de contrastar una hipótesis de partida tomando una decisión estadística basada en el cálculo de probabilidades. En este momento hay que aclarar que el investigador utiliza dos tipos de hipótesis: las sustantivas y las estadísticas (Kerlinger, 1986). La hipótesis sustantiva es la hipótesis de investigación que plantea el investigador, mientras que la hipótesis

estadística —o hipótesis de nulidad— es una afirmación que se realiza sobre uno o más parámetros o sobre la forma de las distribuciones en la población de estudio.

Si por el contrario el objetivo del análisis es sintetizar los datos observados en la realidad con el objeto de detectar estructuras subyacentes, y sin pretender realizar inferencias a otras unidades de análisis distintas de las que se está analizando, entonces recurriremos a la estadística descriptiva. Estas técnicas de análisis descriptivo se utilizan en general como punto de partida en cualquier investigación, para realizar una primera exploración que nos informe del comportamiento de los datos. Este tipo de análisis nos permite, entre otras cosas, detectar la presencia de valores anómalos.

En segundo lugar, el número de variables que se pretenda analizar también repercutirá en la elección de la técnica estadística. En este sentido, el investigador dispone de técnicas estadísticas univariantes —si su interés se centra en el análisis de una sola variable—, técnicas estadísticas bivariantes —si se pretende estudiar la relación entre dos variables—, y técnicas estadísticas multivariantes —si se pretende estudiar la relación entre más de dos variables—.

En tercer lugar, la escala en la que se hayan medido las variables también incidirá en las técnicas estadísticas que se puedan utilizar. Siguiendo la clasificación de Stevens (1946, 1951), las escalas de medida son cuatro: nominal, ordinal, de intervalo y de razón. La escala de medida de una variable determina las transformaciones matemáticas que se pueden realizar con sus valores, por lo que condiciona los índices estadísticos que se pueden aplicar para describir su distribución y las técnicas estadísticas que se pueden aplicar sobre ésta (Michell, 1986).

En cuarto y último lugar, la naturaleza de los datos también es un elemento esencial que repercute en la elección de la técnica estadística. La cuestión es si los datos se han obtenido de los mismos individuos o si, por el contrario, los datos provienen de individuos diferentes. En el primer caso se dice que los datos están relacionados y en el segundo se dice que los datos son independientes. Por lo que se refiere a la estadística inferencial, existen algunas técnicas que se fundamentan en el supuesto de que las observaciones sean independientes, por lo que no deberán utilizarse en ningún caso si se viola este supuesto.

En resumen, el investigador que alcance la fase de análisis de datos deberá contestar a las siguientes cuestiones para seleccionar las técnicas estadísticas que habrá de utilizar:

1. ¿Debo utilizar técnicas de estadística descriptiva o técnicas de estadística inferencial?
2. ¿Debo utilizar técnicas univariantes, bivariantes o multivariantes?
3. ¿Cuál es la escala en que he medido mis variables?
4. Los datos que quiero analizar, ¿son independientes o están relacionados?

Una vez que el investigador haya seleccionado una o varias técnicas estadísticas, procederá a analizar sus datos y deberá interpretar los resultados. Por lo que se refiere a esta interpretación cabe diferenciar entre la interpretación puramente estadística y a la interpretación sustantiva. Tal como indicábamos al iniciar este epígrafe, la interpretación estadística de los resultados se sitúa, dentro del proceso de una investigación, en el nivel analítico-estadístico, mientras que la interpretación sustantiva de los resultados se sitúa en el nivel teórico-conceptual. En efecto, la investigación concluye cuando el investigador regresa a su marco teórico para recuperar su hipótesis sustantiva inicial y confrontarla con los resultados que ha obtenido en el análisis de datos. Como consecuencia de esta confrontación, el investigador confirmará o no su hipótesis original y elaborará sus conclusiones. Keppel (1991) se refiere a esta última etapa de la investigación como *fase reconstructiva*, donde una serie de hechos que no se conocían antes de la realización de la investigación se integran ahora dentro del marco teórico-conceptual.

2.3. TIPOS DE DISEÑOS

Arnau (1995a) define el diseño de investigación como *un plan estructurado de acción que, en función de unos objetivos básicos, está orientado a la obtención de información o datos relevantes a los problemas planteados* (p. 27). Así, el diseño de una investigación se entiende como el plan de actuación que permitirá al investigador recoger los datos para solucionar el problema de su investigación. Para elaborar este plan, el investigador deberá tomar una serie de decisiones. Por ejemplo: decidirá cómo seleccionará la muestra, el número de grupos con los que va a trabajar, la manera en que formará los grupos, el número de observaciones

que se registrará de cada unidad, la forma en que se medirán las variables, los controles que se aplicarán sobre el fenómeno que quiere estudiar, etcétera.

De acuerdo con Arnau (1995a) hay que diferenciar entre tres tipos de diseños: diseños experimentales, diseños cuasi-experimentales y diseños no experimentales. Esta misma terminología es la que proponen Pedhazur y Pedhazur-Schmelkin (1991). No obstante, hay otros autores como Campbell y Stanley (1966) que distinguen entre diseño experimental, cuasi-experimental y correlacional, y otros que reducen las estrategias de investigación a dos: diseños experimentales y diseños *ex post facto* (Dwyer, 1983; Kerlinger, 1986; Roscoe, 1975).

En nuestro caso, vamos a adoptar la clasificación propuesta por Arnau (1995a) y Pedhazur y Pedhazur-Schmelkin (1991). Las diferencias esenciales entre los tres tipos de diseños —experimentales, cuasi-experimentales y no experimentales— ya se han comentado anteriormente y estriban en los principios de manipulación de la variable independiente, el control de las variables de confundido y la aleatorización en la formación de los grupos. Dentro de los diseños experimentales se distingue, a su vez, entre diseños experimentales clásicos y diseños de N=1, cuya diferencia esencial es que en los primeros se trabaja con grupos de individuos mientras que en los segundos se trabaja con una sola unidad de observación.

2.3.1. Diseños experimentales clásicos

2.3.1.1. Definición y características

Comenzaremos este epígrafe ofreciendo algunas definiciones del concepto de diseño experimental.

Kirk (1995) define el término diseño experimental como *un plan para asignar los sujetos a las condiciones experimentales y el análisis estadístico asociado con ese plan* (p. 1). Continúa indicando que el diseño de un experimento para investigar una hipótesis científica o de investigación implica un número de actividades interrelacionadas:

- 1) *Formulación de la hipótesis estadística a partir de la hipótesis científica.*
- 2) *Determinación de las condiciones experimentales que van a utilizarse (variable independiente), la medida que va a registrarse (variable dependiente) y las variables de posible confundido.*

- 3) *Especificación del número de sujetos y de la población de la que se extraerá la muestra.*
- 4) *Especificación del procedimiento para asignar los sujetos a las condiciones experimentales.*
- 5) *Determinación del análisis estadístico que se llevará a cabo.*

Keppel (1991) nos dice que un experimento consiste en *un plan cuidadosamente elaborado y ejecutado para recoger y analizar datos* (p. 4). Continúa diciendo que los requisitos de un experimento son: *(1) distintos tratamientos se administran a distintos grupos de sujetos (o los mismos sujetos en diferentes órdenes) y (2) se observa y se registra la respuesta de los sujetos después de administrar los tratamientos.*

Arnau (1990a) define el diseño experimental como *un plan estructurado de acción tendente a la demostración de relaciones de carácter causal entre la variable independiente y la dependiente. Con este objeto, el diseño experimental recoge los tres supuestos implícitos en el enfoque experimental: (a) aleatorización, (b) control y (c) manipulación.*

De las definiciones anteriores se desprenden las principales características del diseño experimental:

1. Manipulación de la variable independiente: El experimentador interviene de forma directa sobre el fenómeno que quiere estudiar con el objeto de provocar cambios sobre el mismo. Tales cambios se registrarán en la variable dependiente de la investigación. En muchas ocasiones para referirnos a la manipulación de la variable independiente decimos que hemos introducido un tratamiento. Por esta razón algunos autores utilizan el término *variable de tratamiento* y *variable independiente* como sinónimos. La variable independiente o de tratamiento, por definición, tendrá como mínimo dos valores, niveles o tratamientos, y su manipulación implica que el investigador asignará aleatoriamente estos valores a los distintos grupos.
2. Control de las variables de confundido: El experimentador controlará todas las variables que puedan incidir sobre el fenómeno que está estudiando. En un caso ideal, el experimentador debería asegurarse de que la variable dependiente varía *exclusivamente* por su intervención, de manera que debería controlar toda variable que incida sobre ésta. Dicho de otro modo, la

variabilidad en la variable dependiente se debería poder atribuir a la acción de la variable independiente. Como acabamos de decir, éste sería el caso ideal. En la realidad es imposible controlar absolutamente todas las variables de confundido, por lo que el investigador deberá poner todo su empeño en controlar todas las que pueda y minimizar el efecto de aquellas que no pueda controlar completamente.

3. Equivalencia inicial de los grupos: El investigador, cuando opta por un diseño experimental de grupos, asignará aleatoriamente los sujetos o unidades a los grupos experimentales. El objetivo de la formación aleatoria de los grupos es conseguir que éstos sean inicialmente equivalentes. ¿Para qué nos sirve esto? La respuesta es muy clara: si los grupos son equivalentes antes de aplicarles sus respectivos tratamientos, entonces, cualquier diferencia entre ellos después de la intervención se podrá atribuir a la acción del experimentador. Dicho de otro modo: la equivalencia inicial de los grupos garantiza que las diferencias que encontremos entre éstos después de la intervención —y supuesto que se hayan controlando el resto de variables extrañas— se puedan atribuir a la acción de la variable independiente. Obviamente la aleatorización garantizará la equivalencia de los grupos sólo si se forman grupos con un tamaño grande. La lógica subyacente a la aleatorización en la formación de los grupos es que cualquier variable extraña de sujeto quedará repartida aleatoriamente en los distintos grupos, por lo que se neutralizará su efecto.

Estas características del diseño experimental son las que le conducen a presentar una alta validez interna —puesto que es un diseño que nos permitirá establecer relaciones causales inequívocas entre variables—, y una baja validez ecológica —puesto que aborda el fenómeno en una situación tan artificial que hace que se cuestione la posibilidad de generalizar los resultados a una situación natural (recordemos que en un experimento no sólo se provoca el fenómeno sino que, además, se aísla)—. También son estas características las que determinan el ámbito de aplicación más frecuente del diseño experimental. Aunque podemos encontrar algunas excepciones, se trata de un tipo de diseños que se utilizan principalmente en investigaciones de carácter básico que se llevan a cabo en laboratorios. Este es el tipo de contexto que facilita al investigador el máximo control del fenómeno que estudia.

Kerlinger (1986) afirma que el propósito del diseño experimental respecto a la variancia de la variable dependiente es triple y se resume en lo que él mismo denominó el principio del MAX-MIN-CON. Primero: el investigador debe maximizar la *variancia sistemática primaria* de la variable dependiente. Segundo: debe minimizar la *variancia del error* de la variable dependiente. Y, tercero: debe controlar la *variancia sistemática secundaria* de la variable dependiente.

La variancia sistemática se produce cuando los datos (esto es, la variable dependiente) varían en un sentido determinado. Se trata de variabilidad entre-grupos o entre-tratamientos. Esta variabilidad la puede provocar el efecto de la variable independiente —y en ese caso se habla de *variancia sistemática primaria*— o la puede provocar variables extrañas que actúan de forma sistemática sobre la variable dependiente —en cuyo caso, se habla de *variancia sistemática secundaria*—. Dado que la variancia sistemática primaria recoge el principal efecto que se quiere estudiar —recordemos que la intención del investigador es que la variable dependiente varíe por el efecto de la variable independiente—, deberemos maximizarla. De este modo, el efecto de la variable independiente será tan amplio y claro como sea posible. Por el contrario, puesto que la variancia sistemática secundaria recoge el efecto de variables extrañas que provocan variabilidad entre-grupos o entre-tratamientos, deberemos controlarla. La ausencia de control de la variancia sistemática secundaria pone en peligro la relación causal que se pretende establecer.

La variancia de error, por su parte, es variabilidad que presenta la variable dependiente debida a factores aleatorios. Ésta es la razón por la que esta variancia también recibe el nombre de *variancia aleatoria*. Este tipo de variabilidad suele deberse a errores de medida provocados por la utilización de instrumentos mal calibrados o a otros factores relacionados con los propios sujetos, la situación experimental o las condiciones ambientales. Se trata de variabilidad intra-grupo o intra-tratamiento. Dada su naturaleza aleatoria, la variancia de error no se podrá controlar y lo único que estará al alcance del investigador es reducirla o minimizarla. Este objetivo se consigue, en parte, a través de la estructura del diseño.

2.3.1.2. Clasificación de los diseños experimentales clásicos

Vamos a presentar cuatro criterios de clasificación de los diseños experimentales clásicos siguiendo la propuesta de Viader (1995).

El primer criterio clasifica los diseños en función del número de variables independientes que se manipulan. Se distingue, así, entre *diseños simples*, en los que se manipula una sola variable independiente, y *diseños factoriales*, en los que se manipulan dos o más variables independientes. La ventaja del diseño factorial respecto al diseño simple es que nos permite estudiar no sólo el efecto de cada una de las variables independientes por separado —los llamados *efectos principales*—, sino que también nos permite estudiar el efecto conjunto de las mismas —los llamados *efectos interactivos* o de interacción entre las variables—. El efecto interactivo recoge el efecto simultáneo de dos o más variables independientes sobre la variable dependiente. En definitiva, con esta información, el diseño factorial nos permitirá detectar si el efecto de una variable independiente sobre la variable dependiente es distinto en función de con qué valor de la otra variable independiente se combina. En términos técnicos, podremos estudiar los *efectos simples* de cada variable independiente y determinar en qué medida difieren.

No es nuestra intención extendernos en este punto, pero sí queremos señalar que los diseños factoriales pueden presentar una estructura de cruzamiento completa o pueden presentar una estructura incompleta. Un diseño factorial con estructura de cruzamiento completa es aquél en el que los distintos niveles de cada variable independiente se combinan con los distintos niveles de la otra variable independiente, obteniéndose todas las posibles combinaciones de valores. Un diseño factorial incompleto (aquí se incluyen los denominados diseños fraccionados y los diseños anidados o jerárquicos) es aquél en el que no se utilizan todas las posibles combinaciones entre valores de las variables independientes, sino que sólo se utiliza una parte o fracción de éstas. Este tipo de diseños se utiliza cuando se trabaja con muchas variables independientes, por lo que la estructura de cruzamiento completa requeriría muchos grupos experimentales. El problema básico que presenta este tipo de diseños es que no nos permiten estudiar todas las interacciones.

El segundo criterio clasifica los diseños en función del número de variables dependientes que se registren. Así nos encontramos con *diseños univariantes*, en los que registra una sola variable dependiente, y con *diseños multivariantes*, en los que se registra más de una variable dependiente.

El tercer criterio clasifica los diseños en función del número de observaciones por sujeto y condición experimental. Se diferencia así entre *diseño transversal* (también denominado diseño *cross-sectional* o *estático*) y *diseño longitudinal*. En el primero disponemos de una sola medida u observación por sujeto y condición, mientras que

en el segundo disponemos de más de una medida u observación. La diferencia esencial entre uno y otro tipo de diseño radica en el objetivo que persiguen. Si bien el primero se interesa en el estudio de un fenómeno en un momento puntual del tiempo, el segundo, por el contrario, persigue el estudio de la evolución temporal del fenómeno —en definitiva, da respuesta a cuestiones acerca de procesos como la persistencia, el cambio, el crecimiento o el desarrollo (Pedhazur y Pedhazur Schmelkin, 1991)—.

El cuarto criterio de clasificación diferencia los diseños en función de su capacidad para controlar las variables extrañas y reducir la variancia de error. Esta reducción de la variancia de error repercute en que el diseño sea más o menos sensible para detectar el efecto del tratamiento. Así, ordenados de menos a más reducción de la variancia de error, tenemos los *diseños de grupos al azar*, los *diseños de bloques al azar* y los *diseños de medidas repetidas*. Nos vamos a dedicar ahora a presentar con más detalle cada uno de ellos.

Los diseños de grupos al azar se caracterizan porque en ellos la aleatorización interviene a tres niveles (Viader, 1995). Primero: La muestra se selecciona aleatoriamente de la población, con objeto de que todos los individuos de la población tengan la misma probabilidad de pertenecer a la muestra. Segundo: Los grupos se forman al azar, para que cualquier individuo de la muestra tenga la misma probabilidad de pertenecer a uno u otro grupo. Tercero: Los tratamientos o valores de la variable independiente se asignan aleatoriamente a los grupos experimentales, para que todos los grupos tengan la misma probabilidad de recibir uno u otro tratamiento. La principal ventaja de este tipo de diseños ya se ha comentado anteriormente: al formar los grupos al azar, cualquier variable extraña que pudiera afectar a la variable dependiente también quedará repartida al azar en los grupos, por lo que no incidirá de forma sistemática sobre la variable dependiente que interesa al investigador. Como principales desventajas vamos a destacar dos. La primera es que necesitaremos un número amplio de sujetos o unidades que formen parte de cada uno de los grupos para que se alcance la deseada equivalencia inicial entre éstos. La segunda es que presentan una elevada variancia del error, por lo que son diseños poco sensibles para detectar el efecto de los tratamientos.

Los diseños de bloques al azar son aquellos en los que el investigador sabe que hay una variable extraña que puede incidir sobre su variable dependiente, por lo que decide introducirla de forma sistemática en su diseño. Esta variable se

denomina variable de bloqueo. Para ello, formará bloques de unidades experimentales homogéneos respecto a la variable de bloqueo —cada bloque deberá incluir, como mínimo, tantas unidades como grupos experimentales—, y luego las unidades experimentales de cada bloque se asignarán al azar a cada uno de los grupos experimentales. De esta manera, todos los grupos serán homogéneos respecto a la variable de bloqueo. En ocasiones se dice que en este tipo de diseños se lleva a cabo una aleatorización con restricciones. El objetivo de los diseños de bloques es reducir la variancia de error para alcanzar una estimación más precisa del efecto del tratamiento. Para ello, se aísla la variabilidad en la variable dependiente atribuible a la variable extraña o de bloqueo, de manera que ésta no aparezca en las estimaciones del efecto del tratamiento ni del error (Kirk, 1995). Aunque a partir del texto anterior pudiera pensarse que en este tipo de diseños se utiliza una sola variable de bloqueo, en realidad se pueden utilizar más, dando origen a diseños con estructuras más complejas como *el diseño de cuadrado latino* o *el diseño grecolatino*.

La aplicación de la técnica de bloqueo sólo tendrá sentido si se cumplen dos requisitos fundamentales. En primer lugar, debe haber relación entre la variable de bloqueo y la variable dependiente. En segundo lugar, la variable independiente y la variable de bloqueo no deben interactuar. La razón del primer requisito subyace a la propia concepción de la técnica: se supone que la variable de bloqueo, en caso de que no se controle, provocará cambios no deseados en la variable dependiente. La razón del segundo requisito es que si la variable de bloqueo interactúa con la variable independiente, la primera dejará de ser un mero artificio metodológico para reducir la variancia de error y se convertirá en un elemento crucial que condiciona la relación entre el tratamiento y la variable dependiente. En ese caso, el papel de la variable de bloqueo en la investigación pasará de ser una variable extraña que hay que controlar a ser una variable independiente cuyo efecto se debe estudiar.

En un diseño de medidas repetidas el investigador registrará dos o más medidas de la variable dependiente para cada sujeto o unidad experimental. Cada una de estas medidas se tomará bajo la acción de cada una de las condiciones de tratamiento. En consecuencia, algunas de las diferencias que hay entre los diseños de medidas repetidas y los diseños que hemos visto hasta el momento serán las siguientes: (1) En el diseño de medidas repetidas se trabaja con un solo grupo de sujetos, mientras que en el diseño de grupos se trabaja con más de uno —ésta es la razón por la que los primeros reciben el nombre de diseños intrasujeto, y los

segundos reciben el nombre de diseños entresujetos—; y (2) en el diseño de medidas repetidas cada sujeto recibe todos los tratamientos mientras que en el diseño de grupos cada sujeto recibe sólo un tratamiento —en consecuencia, en el primer diseño cada sujeto genera más de un dato, mientras que en el segundo cada sujeto genera un solo dato—.

Maxwell y Delaney (2004) definen el diseño de medidas repetidas o diseño intra-sujeto en un sentido más amplio, incluyendo bajo la misma etiqueta tres situaciones distintas. La primera situación es aquella en la que cada sujeto o unidad de análisis se observa bajo cada condición de tratamiento y se utiliza la misma medida conductual como variable dependiente. Este tipo de diseños se ha denominado en ocasiones diseño *crossover*, especialmente en el ámbito de la investigación médica. La segunda situación es aquella en la que se registran puntuaciones en diferentes tests para cada sujeto. En este caso se dice a veces que se va a realizar un *análisis de perfiles*. La tercera situación es aquella en la que algún aspecto de la conducta de los sujetos se mide en dos o más momentos temporales diferentes. Nótese que esta tercera situación coincide con la que hemos denominado anteriormente diseño longitudinal, en los cuales el objetivo es estudiar el desarrollo de la conducta a lo largo del tiempo.

De acuerdo con Maxwell y Delaney (2004), los diseños de medidas repetidas presentan dos claras ventajas si los comparamos con los diseños de grupos. De una parte, en un diseño de medidas repetidas se obtendrá más información de cada sujeto, puesto que cada uno de ellos nos aportará como mínimo dos medidas. En un diseño de grupos, por el contrario, cada sujeto nos facilitará una sola medida de la variable dependiente. Esto tiene una implicación importante: el número de sujetos que se necesitará en un diseño de medidas repetidas, para alcanzar un determinado nivel de potencia estadística, será inferior al número de sujetos que se necesitará en un diseño de grupos. Cuando el coste para conseguir los sujetos es alto —en términos de dinero, tiempo o esfuerzo—, esta ventaja de los diseños de medidas repetidas será un elemento muy importante que el investigador deberá considerar.

De otra parte, los diseños de medidas repetidas presentan la ventaja de reducir la variancia de error. Puesto que las comparaciones en el diseño de medidas repetidas se realizan intra-sujeto, la variabilidad provocada por las diferencias individuales se elimina del término de error. Es por esto por lo que en este tipo de diseños se dice que el sujeto actúa como su propio control reduciendo una parte de la variancia de error provocada por variables extrañas de efectos aleatorios.

El diseño de medidas repetidas también presenta una serie de inconvenientes, que incluso pueden llegar a hacer que sea imposible la utilización de este diseño. El problema esencial estriba precisamente en su estructura, es decir, en el hecho de que se tome más de una medida para cada sujeto. La situación más clara en la que será imposible recurrir a un diseño de medidas repetidas es cuando el efecto de las condiciones de tratamiento sea irreversible.

Usualmente se habla de dos inconvenientes que pueden darse en el diseño de medidas repetidas: los efectos residuales y los efectos de orden. No obstante, unos y otros se pueden controlar. Los efectos residuales se producen cuando el efecto del tratamiento se prolonga en el tiempo, de manera que este tratamiento podría seguir actuando sobre la variable dependiente cuando se introduce el siguiente tratamiento. El efecto de confundido aquí es evidente. La solución que se propone en este caso es, primero, conocer la latencia del efecto de los tratamientos, y, segundo, incrementar el intervalo de tiempo entre las sucesivas administraciones de los distintos tratamientos. Esta es la manera de asegurarse de que cuando se administra un tratamiento el efecto del tratamiento anterior sobre la variable dependiente ya ha desaparecido. Los efectos de orden, por su parte, se pueden producir si todas las condiciones de tratamiento se administran en el mismo orden. Así, por ejemplo, pueden introducirse en la investigación variables extrañas de efectos sistemáticos como la fatiga o el aprendizaje. Para controlar el efecto de estas variables el investigador dispone de dos alternativas. Una es que cada sujeto reciba las condiciones de tratamiento en un orden distinto —lo que se conoce en términos metodológicos como técnica de contrabalanceo—, y otra es que cada sujeto reciba las condiciones de tratamiento en un orden aleatorio.

Un caso especial de diseño es aquél en el que se combina un diseño de grupos con un diseño de medidas repetidas. Este tipo de diseño recibe distintas denominaciones: diseño factorial mixto o diseño de medidas parcialmente repetidas. Se trata de un diseño en el que trabajaremos con dos o más grupos de sujetos, de los cuales tomaremos dos o más medidas de la variable dependiente.

2.3.1.3. Análisis de datos en el diseño experimental clásico

En este apartado no pretendemos profundizar en las distintas técnicas estadísticas que un investigador tendrá a su disposición para analizar los datos de su experimento. Tales técnicas son muchas, por lo que presentarlas todas se escapa

del objetivo de este texto. No obstante, consideramos que sí debemos presentar la técnica estadística que se utiliza con más frecuencia para el análisis de datos que proceden de experimentos. A esta técnica vamos a dedicar las siguientes páginas. Se trata del análisis de la variancia, conocida con el acrónimo AVAR o ANOVA.

Antes de comenzar queremos subrayar que el hecho de que el AVAR se utilice con mucha frecuencia en el análisis de experimentos no quiere decir que esta técnica de análisis sea exclusiva de este tipo de diseños. Por el contrario, se utilizará tanto en diseños de tipo cuasi-experimental como en diseños de tipo no experimental.

El AVAR es una técnica estadística que nos permite estudiar la relación entre una variable de agrupación y una variable cuantitativa —en términos experimentales, nos permitirá estudiar el efecto de la variable independiente sobre la variable dependiente—. La variable cuantitativa necesariamente ha de estar medida, como mínimo, en una escala de intervalo. En este sentido, se trata de una técnica que estudia un tipo de efecto similar a la prueba *t* de Student, no obstante, a diferencia de ésta, nos abre la oportunidad de estudiar el efecto de variables de agrupación con más de dos valores.

De hecho, y a pesar de que por su nombre pudiera dar lugar a confusión, el AVAR —igual que la prueba *t* de Student— es una técnica que permite contrastar medias, de manera que la hipótesis nula de la que parte es la siguiente:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

Esta hipótesis nos indica que todas las medias poblacionales son idénticas. Por su parte, la hipótesis alternativa se expresa como sigue:

$$H_1 : \exists i, j \mid \mu_i \neq \mu_j$$

Lo cual quiere decir que, por lo menos, hay dos medias poblacionales que difieren.

A continuación vamos detenernos para presentar los principios básicos de tres tipos de AVAR. Primero: el AVAR para un diseño simple de grupos al azar. Segundo: el AVAR para un diseño factorial de grupos al azar. Y, tercero: el AVAR para un diseño simple de medidas repetidas. Nos centraremos en estos tres tipos de AVAR, porque nos permitirán: (1) destacar las diferencias entre un AVAR simple y un AVAR factorial, lo que quiere decir que veremos cómo el AVAR nos permite estudiar tanto

el efecto de una sola variable de agrupación —el AVAR simple o unifactorial— como el efecto de más de una variable de agrupación —el AVAR factorial—, y (2) destacar las diferencias entre un AVAR de grupos y un AVAR de medidas repetidas, o, lo que es lo mismo, entre un AVAR para datos independientes y un AVAR para datos relacionados.

AVAR UNIFACTORIAL PARA MUESTRAS INDEPENDIENTES

Este tipo de análisis podrá utilizarse, por ejemplo, para el análisis de los datos de un diseño simple de grupos al azar. Esto es: para un diseño en el que se manipula una sola variable independiente y se trabaja con más de un grupo. En la Tabla 2.3 se muestra la estructura de la tabla de datos de partida para este tipo de AVAR.

Tabla 2.3. Matriz de datos para un AVAR unifactorial de muestras independientes.

	Grupo 1	Grupo 2	...	Grupo k	Total
Individuo 1	y_{11}	y_{12}	...	y_{1k}	
Individuo 2	y_{21}	y_{22}	...	y_{2k}	
·	·	·		·	
·	·	·		·	
·	·	·		·	
Individuo n	y_{n1}	y_{n2}	...	y_{nk}	
Sumas	$\sum_{i=1}^n y_{i1}$	$\sum_{i=1}^n y_{i2}$...	$\sum_{i=1}^n y_{ik}$	$\sum_{j=1}^k \sum_{i=1}^n y_{ij}$
Medias	$\bar{y}_{.1}$	$\bar{y}_{.2}$...	$\bar{y}_{.k}$	$\bar{y}_{..}$
Nº individuos	n_1	n_2	...	n_k	n

El modelo estadístico subyacente al AVAR es el siguiente:

$$y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

donde y_{ij} es la puntuación del individuo i del grupo j , μ es la media poblacional bajo el supuesto de hipótesis nula verdadera, α_j es el efecto del valor de la variable de agrupación que define el grupo j y ε_{ij} es el residual. Este residual incluye tanto el efecto de otras variables como el error de medición.

El modelo anterior puede transformarse en esta expresión:

$$y_{ij} - \mu = \alpha_j + \varepsilon_{ij}$$

Podemos utilizar como estimadores para μ y α_j , los siguientes valores conocidos:

$$\hat{\mu} = \bar{y}_{..}$$

$$\hat{\alpha}_j = \bar{y}_{.j} - \bar{y}_{..}$$

Por lo que se refiere a ε_{ij} lo sustituimos por la siguiente expresión:

$$y_{ij} - \bar{y}_{.j}$$

A partir de aquí el desarrollo del modelo del AVAR es el siguiente:

$$y_{ij} - \bar{y}_{..} = (\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{.j})$$

Si elevamos al cuadrado e incluimos todos los individuos, la expresión anterior se transforma en la siguiente:

$$\sum_{j=1}^k \sum_{i=1}^n (y_{ij} - \bar{y}_{..})^2 = \sum_{j=1}^k \sum_{i=1}^n [(\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{.j})]^2$$

Entonces,

$$\begin{aligned} \sum_{j=1}^k \sum_{i=1}^n (y_{ij} - \bar{y}_{..})^2 &= \sum_{j=1}^k \sum_{i=1}^n [(\bar{y}_{.j} - \bar{y}_{..})^2 + (y_{ij} - \bar{y}_{.j})^2 + 2(\bar{y}_{.j} - \bar{y}_{..})(y_{ij} - \bar{y}_{.j})] \\ &= \sum_{j=1}^k \sum_{i=1}^n (\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{j=1}^k \sum_{i=1}^n (y_{ij} - \bar{y}_{.j})^2 + 2 \sum_{j=1}^k \sum_{i=1}^n (\bar{y}_{.j} - \bar{y}_{..})(y_{ij} - \bar{y}_{.j}) \end{aligned}$$

La demostración de que el último término de la expresión anterior es igual a cero se presenta a continuación:

$$2 \sum_{j=1}^k \sum_{i=1}^n (\bar{y}_{.j} - \bar{y}_{..})(y_{ij} - \bar{y}_{.j}) = 0$$

Para cualquier valor de j en el primer sumatorio de la anterior expresión es cierta la siguiente igualdad,

$$\sum_{i=1}^n (\bar{y}_{.j} - \bar{y}_{..})(y_{ij} - \bar{y}_{.j}) = (\bar{y}_{.j} - \bar{y}_{..}) \sum_{i=1}^n (y_{ij} - \bar{y}_{.j})$$

Eso es así porque fijado j en el primer sumatorio la diferencia $(\bar{y}_{.j} - \bar{y}_{..})$ es una constante. En consecuencia, por medio de la propiedad distributiva del producto respecto a la suma (o sea, sacar factor común), es cierta la anterior expresión. Por tanto,

$$2 \sum_{j=1}^k \sum_{i=1}^n (\bar{y}_{.j} - \bar{y}_{..})(y_{ij} - \bar{y}_{.j}) = 2 \sum_{j=1}^k \left[(\bar{y}_{.j} - \bar{y}_{..}) \sum_{i=1}^n (y_{ij} - \bar{y}_{.j}) \right]$$

Como para todo valor de j será cierto que $\sum_{i=1}^n (y_{ij} - \bar{y}_{.j}) = 0$ se concluye que,

$$2 \sum_{j=1}^k \sum_{i=1}^n (\bar{y}_{.j} - \bar{y}_{..})(y_{ij} - \bar{y}_{.j}) = 0$$

De este modo, la expresión final es la siguiente:

$$\sum_{j=1}^k \sum_{i=1}^n (y_{ij} - \bar{y}_{..})^2 = \sum_{j=1}^k \sum_{i=1}^n (\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{j=1}^k \sum_{i=1}^n (y_{ij} - \bar{y}_{.j})^2$$

Así, la variación total queda dividida en dos partes: la primera, corresponde a la variación entregrupos (sumas de cuadrados de las distancias entre las medias de los grupos y la media global); la segunda, a la variación intragrupo (sumas de cuadrados de las distancias entre las puntuaciones de los individuos y la media del grupo al que pertenecen).

Si dividimos las sumas de cuadrados entre los grados de libertad obtendremos variancias, de manera que con la anterior expresión logramos dos estimaciones independientes de la variancia poblacional bajo el supuesto de hipótesis nula. Por una parte, la variancia entregrupos, a la que también se denomina variancia explicada, puesto que recoge la variancia que puede asignarse al efecto de la variable de agrupación. Por otra parte, la variancia intragrupo, también conocida

como variancia no explicada, puesto que recoge fundamentalmente variabilidad aleatoria.

Los grados de libertad se calculan como sigue:

$n-1$ para la variancia total, siendo n el tamaño de la muestra
 $k-1$ para la variancia entregrupos, siendo k el número de grupos
 $n-k$ para la variancia intragrupo

Porque tenemos dos estimaciones independientes de la variancia poblacional, podemos calcular el valor de F , que se obtiene dividiendo la variancia entregrupos entre la variancia intragrupo.

$$F = \frac{S^2_{\text{entregrupos}}}{S^2_{\text{intragrupo}}}$$

Si ambas variancias son iguales, entonces el valor de F será 1, lo que nos llevará a concluir que no hay diferencias entre los grupos.

Siempre que F sea mayor que 1, por el contrario, diremos que hay diferencias entre los grupos. No obstante, para decidir si dichas diferencias son o no estadísticamente significativas habremos de obtener el nivel de significación del estadístico F . Para conocer el nivel de significación será preciso obtener la Prob ($F \geq F_{v_1, v_2}$), siendo $v_1 = k-1$ y $v_2 = n-k$

Dado que el AVAR es un contraste global, siempre que concluyamos que hay diferencias estadísticamente significativas entre las medias, será preciso llevar a cabo *contrastos parciales* entre las medias de los grupos. Tales contrastes nos permitirán concretar las diferencias detectadas por el AVAR (véanse las distintas técnicas estadísticas para comparaciones múltiples de medias independientes, por ejemplo, en Hays, 1994 y Ramsey, 1993).

El AVAR suele presentarse en forma de tabla, siguiendo la estructura que se muestra en la Tabla 2.4:

Tabla 2.4. Cuadro resumen del AVAR unifactorial para datos independientes.

Fuentes de Variación	Suma de cuadrados (SC)	Grados de libertad	Variancia	F
Factor entregrupos	$SC_{entre} = \sum_{j=1}^k \sum_{i=1}^n (\bar{y}_{.j} - \bar{y}_{..})^2$	$k-1$	$S_{entre}^2 = \frac{SC_{entre}}{k-1}$	$F = \frac{S_{entre}^2}{S_{intra}^2}$
Factor intragrupo	$SC_{intra} = \sum_{j=1}^k \sum_{i=1}^n (\bar{y}_{ij} - \bar{y}_{.j})^2$	$n-k$	$S_{intra}^2 = \frac{SC_{intra}}{n-k}$	
Total	$SC_{total} = \sum_{j=1}^k \sum_{i=1}^n (y_{ij} - \bar{y}_{..})^2$	$n-1$		

Dado que el estadístico F sólo nos informa acerca de si hay o no diferencias estadísticamente significativas entre las muestras, pero no sobre la magnitud de esa diferencia, se han desarrollado varios índices que nos facilitan información sobre el tamaño del efecto de las variables (Cohen, 1988). De hecho, cada vez son más las revistas científicas que indican en su línea editorial que no sólo se debe informar del valor exacto del nivel de significación o p valor, sino también del tamaño del efecto (American Psychological Association, 1994; Wilkinson & APA Task Force on Statistical Inference, 1999). Uno de los índices más utilizado es η^2 .

$$\hat{\eta}^2 = \frac{SC_{entregrupos}}{SC_{total}}$$

Este índice, que oscila entre 0 y 1, se interpreta como la proporción de variabilidad de la variable cuantitativa (generalmente la variable dependiente) atribuible a la variable de agrupación (generalmente la variable independiente). En el caso de que nos encontremos ante un AVAR multifactorial el tamaño del efecto puede calcularse tanto para los efectos principales como para los efectos interactivos.

Existen tres supuestos que han de cumplirse si queremos aplicar un AVAR:

1. Independencia de las observaciones: se refiere a que las puntuaciones de los distintos individuos no han de covariar entre sí. La selección aleatoria de las muestras y la asignación aleatoria de cada individuo a cada grupo nos permitirá alcanzar con bastante seguridad este supuesto. Esta condición es necesaria para utilizar el modelo que descompone la puntuación de un sujeto de la siguiente manera:

$$y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

2. Normalidad de los datos: el conjunto de residuales en la población debe distribuirse según una ley normal. Existen distintas pruebas para analizar la normalidad de una distribución, entre otras, la prueba de Shapiro-Wilk o la prueba de Kolmogorow-Smirnov. No obstante, es preciso mencionar que el AVAR es bastante robusto ante pequeñas desviaciones de la normalidad —en concreto, para ligeras asimetrías—. Este supuesto debe cumplirse para poder utilizar la razón de variancias como contraste estadístico.
3. Homocedasticidad: para una correcta utilización del AVAR es necesario que las variancias intragrupo sean homogéneas. Esta condición debe cumplirse para poder realizar una estimación conjunta de la variancia residual. Si queremos comprobar esta condición de aplicación podemos recurrir, por ejemplo, a las pruebas de Hartley, Cochran, Bartlett-Box o Levene. Las dos primeras podrán utilizarse cuando todos los grupos tengan el mismo número de individuos. Por el contrario, la prueba de Bartlett-Box y la de Levene se aplicarán tanto si trabajamos con grupos de igual tamaño como si trabajamos con grupos de diferente tamaño.

AVAR MULTIFACTORIAL PARA MUESTRAS INDEPENDIENTES

En el epígrafe anterior hemos visto cómo podemos utilizar el AVAR unifactorial cuando deseamos estudiar la relación entre una variable de agrupación y una variable cuantitativa. Ahora, presentaremos otra modalidad del AVAR que nos permite analizar la relación entre dos o más variables de agrupación y una variable cuantitativa. En este caso utilizaremos un AVAR multifactorial.

La principal ventaja de la realización del AVAR multifactorial es que, además de poder analizar el efecto de cada una de las variables de agrupación por separado

(también conocido como *efectos principales* de las variables), nos abre la posibilidad de analizar la interacción (o *efecto interactivo*). El análisis de la interacción nos permitirá detectar si el efecto de uno o más valores de la variable de agrupación está condicionado a alguno o algunos de los valores de la otra variable de agrupación.

A modo de ejemplo presentamos en la Tabla 2.5 la matriz de datos para el caso más sencillo del AVAR multifactorial para muestras independientes. Se trata del caso 2 x 2, donde sólo intervienen dos variables de agrupación, cada una de ellas con sólo dos valores.

Tabla 2.5. Matriz de datos para un AVAR multifactorial 2 x 2 de muestras independientes

		VARIABLE B			
		b ₁	b ₂	medias	
V A R I A B L E A	a ₁	individuo 1	y ₁₁₁	y ₁₂₁	
		
		individuo n	y _{11n}	y _{12n}	
	medias		$\bar{y}_{11.}$	$\bar{y}_{12.}$	$\bar{y}_{1..}$
	a ₂	individuo 1	y ₂₁₁	y ₂₂₁	
		
individuo n		y _{21n}	y _{22n}		
medias		$\bar{y}_{21.}$	$\bar{y}_{22.}$	$\bar{y}_{2..}$	
medias		$\bar{y}_{.1.}$	$\bar{y}_{.2.}$	$\bar{y}_{...}$	

Para el caso anterior el modelo subyacente al AVAR es el siguiente:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$$

donde y_{ijk} es la puntuación de un individuo, μ es la media poblacional bajo el supuesto de hipótesis nula verdadera, α_i es el efecto del valor i de la variable A, β_j es el efecto del valor j de la variable B, $\alpha\beta_{jk}$ es el efecto de la interacción (o efecto conjunto) de las dos variables y ε_{ijk} es el residual.

El desarrollo de este modelo es similar al que hemos visto al presentar el caso unifactorial, de manera que la variación total, de nuevo, se descompone en la variación explicada y la variación no explicada. En este caso, dentro de la variación explicada se incluye los efectos principales de cada variable (α_i y β_j) y el efecto interactivo ($\alpha\beta_{jk}$). Por lo que se refiere a la variación no explicada, ésta se recoge en el residual o término de error (ϵ_{ijk}).

Obtendremos la variancia para cada uno de los efectos, resultado de dividir su correspondiente suma de cuadrados entre los grados de libertad. En este caso, los grados de libertad se calculan como sigue:

- $a-1$ para la variable A, siendo a el número de valores de la variable A
- $b-1$ para la variable B, siendo b el número de valores de la variable B
- $(a-1)(b-1)$ para la interacción
- $n-ab$ para la residual, siendo n el tamaño de la muestra.

Finalmente, obtendremos una razón F para cada uno de los efectos principales y para la interacción, dividiendo la variancia de cada efecto entre la variancia residual.

Puesto que el AVAR es un contraste global, siempre que nos indique que hay diferencias estadísticamente significativas para alguno de los efectos analizados, deberemos llevar a cabo *contrastos parciales* a fin de estudiar con más detalle tales diferencias. En el caso de que la interacción alcance significación estadística, ésta se estudiará con más detalle mediante el cálculo de los efectos simples. En la Figura 2.1 se ilustran gráficamente algunas posibles interacciones entre variables: los tres casos superiores en los que no se produce interacción y los dos casos inferiores en los que sí hay interacción.

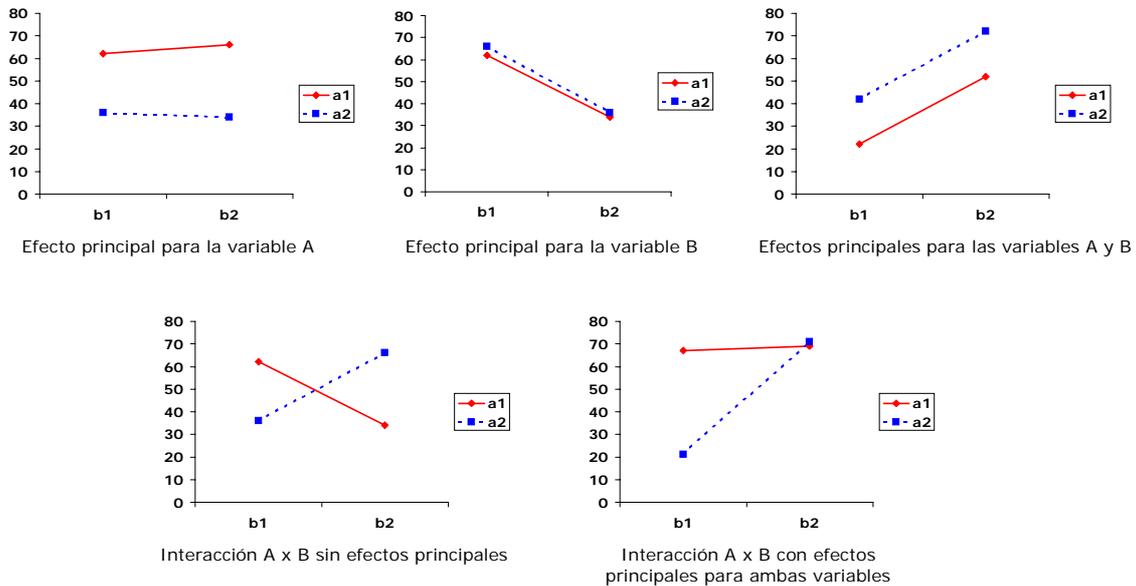


Figura 2.1. Ejemplos gráficos de interacciones para un diseño factorial A x B.

AVAR UNIFACTORIAL PARA MUESTRAS RELACIONADAS

Utilizaremos el AVAR para muestras relacionadas cuando tengamos medidas de los mismos individuos bajo los distintos niveles de la variable que pretendamos analizar (en ocasiones se utiliza el término *tratamiento* para referirse a cada uno de estos niveles). En este caso, nos encontramos con una ventaja respecto al AVAR para muestras independientes: puesto que tenemos varias medidas para cada individuo, la variabilidad debida a diferencias individuales podrá extraerse del término de error, reduciendo así el término residual del análisis respecto al caso del AVAR de muestras independientes.

Si bien el AVAR para muestras relacionadas ofrece la ventaja comentada en el punto anterior con relación al AVAR para muestras independientes, también presenta respecto a éste un inconveniente: se trata del problema de la no independencia de los errores. Dado que el AVAR no es una técnica robusta ante el incumplimiento de este supuesto necesitaremos emplear procedimientos específicos para realizar el análisis en el caso de que se dé dependencia entre las

observaciones. Estos procedimientos se comentarán más adelante en el apartado correspondiente a las condiciones de aplicación.

En la Tabla 2.6 se muestra la estructura de la matriz de datos del AVAR unifactorial para muestras relacionadas.

Tabla 2.6. Matriz de datos para un AVAR unifactorial para muestras relacionadas

	VARIABLE A				
	a ₁	a ₂	...	a _k	medias
individuo 1	y ₁₁	y ₁₂	...	y _{1k}	$\bar{y}_{1.}$
individuo 2	y ₂₁	y ₂₂	...	y _{2k}	$\bar{y}_{2.}$
.
.
.
individuo n	y _{n1}	y _{n2}	...	y _{nk}	$\bar{y}_{n.}$
medias	$\bar{y}_{.1}$	$\bar{y}_{.2}$...	$\bar{y}_{.k}$	$\bar{y}_{..}$

El AVAR unifactorial para muestras relacionadas puede plantearse mediante dos modelos estructurales distintos: uno, denominado *modelo aditivo*, que supone que no hay interacción entre los individuos y la variable que pretendemos analizar; otro, denominado *modelo no aditivo*, que supone que sí que hay interacción. Los modelos estructurales para cada caso son los siguientes:

Modelo aditivo:
$$y_{ij} = \mu + \alpha_j + \pi_i + \varepsilon_{ij}$$

Modelo no aditivo:
$$y_{ij} = \mu + \alpha_j + \pi_i + \pi\alpha_{ij} + \varepsilon_{ij}$$

donde y_{ij} es la puntuación del individuo i en el valor j de la variable, μ es la media poblacional bajo el supuesto de hipótesis nula verdadera, α_j es el efecto del valor j de la variable, π_i es el efecto del individuo i , $\pi\alpha_{ij}$ es el efecto de la interacción (o efecto conjunto) del valor j de la variable y el individuo i , y ε_{ij} es el residual.

Tanto en un caso como en el otro el término de contraste para obtener el valor del estadístico F será el componente residual.

Además de los supuestos que han de cumplirse para la aplicación de cualquier AVAR, en el caso de que estemos trabajando con muestras relacionadas se añade el supuesto llamado de *esfericidad*. Esta condición supone la homogeneidad de las variancias de las diferencias entre pares de tratamientos. El procedimiento más

usual para comprobar el cumplimiento de la condición de esfericidad es el propuesto por Mauchly (1940). El cumplimiento de esta condición es necesario para que el cociente de la variancia debida a los tratamientos (es decir, la variancia explicada por la variable cuyo efecto pretendemos analizar) y la variancia residual siga una distribución F con $(a-1)$ y $(a-1)(n-1)$ grados de libertad. En caso de incumplimiento de esta condición de aplicación el cociente F estará sobreestimado, incrementado así la probabilidad de cometer error tipo I.

Siempre que el modelo sea aditivo se cumplirá la condición de esfericidad. No obstante, si nos encontramos ante un modelo no aditivo (es decir, cuando se produzca interacción entre los individuos y la variable cuyo efecto pretendamos analizar) se incumplirá esta condición.

Puesto que el AVAR no es robusto al incumplimiento de esta condición de aplicación se han propuesto dos alternativas que pueden adoptarse en caso de que se detecte su incumplimiento. Por una parte, Geisser y Greenhouse (1958) y Huynh y Feldt (1976) propusieron aplicar una corrección a los grados de libertad en el AVAR para disminuir la probabilidad de cometer error tipo I. Por otra parte, podemos analizar este tipo de datos mediante el análisis multivariante de la variancia (MANOVA). La ventaja que tiene la utilización del MANOVA es que no requiere el cumplimiento de la condición de esfericidad. En Maxwell y Delaney, (1990), Vasey y Thayer (1987), y Keselman, Keselman y Lix (1995) se ofrece una valoración comparativa entre el AVAR con corrección de grados de libertad y el MANOVA para el análisis de datos de medidas repetidas.

2.3.2. Diseños experimentales de N=1

2.3.2.1. Definición y características

El diseño experimental de N=1 ha recibido otras denominaciones. Señala Arnau (1995b) que este diseño se ha etiquetado como *diseño experimental de caso único* (Barlow y Hersen, 1984; Franklin, Alison y Gorman, 1996; Hersen y Barlow, 1976; Kazdin, 1982; Kennedy, 2005; Shapiro, 1961, 1966), *diseño de replicación intrasujeto* (Arnau, 1984; Gentile, Roden y Klein, 1972), *diseño operante* (Sidman, 1960), *diseño intensivo* (Chassan, 1967) y *diseño de control propio* (Millon y Diefenhaus, 1972). Otros autores lo han denominado *diseño de un solo sujeto* (Bordens y Abbot, 1988; Cozby, 1993) y *diseño conductual* (Balluerka, 2001). Por lo

que refiere a la expresión que adoptaremos aquí, el diseño experimental de $N=1$ recibe este nombre en contraste con el diseño experimental clásico o fisheriano (diseño de $N>1$), al cual hemos dedicado las páginas anteriores. Ato y Vallejo (2007) exponen las diferencias esenciales entre el diseño de $N>1$ y el diseño de $N=1$, indicando que en los primeros se sigue un enfoque nomotético mientras que en los segundos se sigue un enfoque idiográfico. Estos autores nos dicen lo siguiente:

Los investigadores que utilizan el enfoque nomotético exponen dos o más grupos de sujetos a las diferentes condiciones de una o más variables independientes y, tras recoger las puntuaciones promediadas de los distintos miembros de cada grupo, se comparan las varianzas intergrupales con la intragrupal para determinar la probabilidad que tienen las diferencias obtenidas de acontecer por azar. Por el contrario, los investigadores que utilizan el enfoque idiográfico no se centran en la variabilidad existente dentro de la muestra, ni en los promedios grupales; más bien, las respuestas de cada individuo son cuidadosamente examinadas a través de un período relativamente prolongado de tiempo (Ato y Vallejo, 2007, p. 545).

Más adelante, Ato y Vallejo (2007) definen los diseños de $N=1$ como *diseños en los que una misma unidad de observación (sujeto, paciente, participante, organismo o cualquier otra entidad) se mide repetidamente a lo largo del tiempo bajo todos los niveles de, al menos, una variable independiente (p. 546).*

Kennedy (2005) define el diseño de caso único como *aquél que demuestra el control experimental —es decir, la demostración de que una intervención concreta produce un cambio en la conducta— con un único individuo que actúa como participante control y como participante experimental (p. 12)*. En otras palabras, en el diseño de caso único —o de $N=1$ — se toman medidas de la variable de respuesta de forma repetida en una fase de control y en una fase experimental para un único individuo.

Kazdin (2001) dice que *la razón lógica de los diseños experimentales de caso único es similar a la más conocida de los diseños en grupo. Todos los experimentos comparan los efectos de las diferentes condiciones (variables independientes) sobre el comportamiento (p. 207)*. En la experimentación tradicional en grupo, el efecto

de la intervención se evalúa comparando el desempeño de los diferentes grupos experimentales, mientras que en los diseños de N=1 ese efecto se evalúa comparando condiciones diferentes que se presentaron al mismo sujeto a lo largo del tiempo.

Arnau (1995) afirma que los diseños de N=1 *sirven para estudiar los efectos de las variables de tratamiento o intervención sobre la conducta (objeto de estudio) de individuos o grupos de individuos* (p.180). Añade que *cabe destacar dos aspectos básicos es esta clase de estrategia de investigación: la aplicación secuencial y, en muchos casos, retirada de la variable independiente (tratamiento o intervención), y el uso de medidas frecuentes y repetidas del resultado* (p. 180).

A partir de las definiciones anteriores podemos enumerar las características que presentan los diseños de N=1. Son las siguientes:

1. La conducta se estudia a lo largo del tiempo, por lo que se toman registros de ésta en muchos momentos temporales. Esta conducta se conoce como variable de respuesta o resultado (en ocasiones, también se la denomina variable dependiente).
2. La conducta se registra en fases distintas: fases sin intervención —o con ausencia de tratamiento— y fases con intervención —o con administración de tratamiento—. Es, precisamente, la comparación entre fases la que permite evaluar si el tratamiento modifica la conducta.
3. A diferencia de los diseños de grupos, los diseños de N=1 se interesan en el estudio del individuo, por lo que trabajarán con un único sujeto. Se entiende aquí el término *sujeto* en un sentido amplio, puesto que puede tratarse de un solo individuo, paciente u organismo, o de cualquier otra entidad formada por varios individuos. Tanto si se trabaja con un solo individuo como si se trabaja con una entidad formada por varios individuos, el interés en los diseños de N=1 se centra en la variabilidad individual y no en la variabilidad grupal. En definitiva, este tipo de diseño centra su atención en el caso individual y no en el individuo promedio.
4. Puesto que se trata de un diseño experimental, el diseño de N=1 compartirá con los diseños experimentales clásicos dos características esenciales. A saber: la manipulación de la variable independiente y el control de las variables extrañas. Por lo que se refiere a la primera característica, la manipulación implica que el investigador intervendrá de forma directa sobre

el fenómeno que pretenda estudiar. Tal intervención puede consistir en la aplicación de un determinado tratamiento o programa. Por lo que se refiere a la segunda característica —el control de las variables extrañas—, el diseño de N=1 puede recurrir en la recogida de datos a tres controles esenciales: la *línea base*, la *reversión* y la *replicación* (existen otros controles que se explicarán más adelante cuando presentemos la clasificación de estos diseños). La línea base consiste en el registro continuado de la conducta de interés antes de la intervención, y juega en la investigación un papel semejante al del grupo control en un diseño de grupos. Tanto la línea base como el grupo control nos facilitan unas medidas de la variable dependiente en ausencia de intervención, medidas que más tarde se compararán con la fase de intervención, en el diseño de N=1, y con las medidas del grupo experimental, en el diseño de grupos. Sidman (1960) nos indica que la línea base debe presentar tres propiedades: (1) estabilidad, es decir, la conducta no debe variar de una observación a la siguiente; (2) sensibilidad, es decir, debe captar cualquier cambio que se efectúe en la variable independiente, por pequeño que éste sea; y (3) control interno, es decir, la línea base ha de permitir controlar o eliminar los procesos conductuales no deseados. La reversión —también llamada *fase de reversibilidad conductual* o *crítica*— consiste en la incorporación de una fase de registro en la que se retira el tratamiento después de la fase de intervención. Cabe esperar que si el tratamiento es eficaz, la conducta objeto de estudio vuelva al nivel basal. El problema radica en que, en ocasiones, no será posible introducir una fase de retirada de tratamiento, bien sea porque la conducta sea irreversible, bien sea porque haya consideraciones éticas que lo impidan. Por último, la replicación consiste en la incorporación de una nueva fase de registro en la que el tratamiento se vuelve a aplicar después de la fase de retirada de tratamiento. Cabe esperar que si el tratamiento es eficaz se volverá a observar un cambio en la conducta objeto de estudio en el sentido esperado.

5. De lo anterior se desprende que el diseño experimental de N=1, igual que el diseño experimental clásico, tiene como objetivo el establecimiento de relaciones causales inequívocas.
6. A diferencia del diseño experimental clásico, el diseño experimental de N=1 se utiliza con mucha frecuencia en investigación aplicada, fundamentalmente en los ámbitos clínico y educativo.

7. El diseño experimental de N=1 es más flexible y dinámico que el diseño experimental clásico (Kennedy, 2005; Kratochwill, Mott y Dodson, 1984; Poling, Methot y LeSage, 1995). Esto se debe a que, en los primeros, los datos se van recogiendo de forma continua hasta que se completa el experimento. Cuando un investigador utiliza un diseño experimental clásico tiene un plan de recogida de datos cerrado, en el que no podrá introducir modificaciones sin poner en peligro la validez interna de su investigación. En un experimento de N=1, por el contrario, aunque el investigador también tenga un plan de recogida de datos, éste no está completamente cerrado, por lo que puede introducir pequeñas modificaciones. Por ejemplo, el investigador al comenzar su experimento no sabrá cuántas observaciones necesitará tomar de la variable de respuesta en la línea base, por lo que deberá ir recopilando datos hasta que ésta alcance la estabilidad. Lo mismo puede decirse para las fases de tratamiento. Ésta es la razón por la que, como señala Kennedy, el proceso de análisis de la conducta en un diseño de N=1 se resume en la expresión *follow your data* (p. 191).

Franklin, Allison y Gorman (1996) manifiestan que el diseño de N=1 es preferible a los diseños de grupos en tres situaciones:

1. Las becas de investigación son difíciles de conseguir para los profesionales que trabajan en una consulta privada o en una pequeña clínica. Los diseños de N=1 brindan la oportunidad a estos profesionales de contribuir a la ciencia con sus descubrimientos estudiando a pacientes individuales. Su trabajo puede repercutir en beneficios para sus propios pacientes con un coste inferior al de la mayoría de los diseños de grupos. Dado su bajo coste, los diseños de N=1 pueden resultar también útiles en las primeras fases de una investigación como un medio para generar datos piloto.
2. Los diseños de N=1 también son muy útiles cuando se quiere estudiar el proceso de cambio de forma intensiva. Al tomar muchas medidas de la variable de respuesta a lo largo del tiempo, este tipo de diseño permite al investigador observar los aspectos dinámicos del cambio.
3. Por último, la mayoría de los trabajos que se desarrollan en el ámbito clínico pretenden responder a la cuestión acerca de si un determinado tratamiento funciona para un paciente concreto. El interés de estos trabajos, pues, se

centra en explicar el comportamiento de un solo individuo y no en el comportamiento del individuo promedio (Sidman, 1952). La primera cuestión nos la permitirá responder un diseño de $N=1$ —sólo los diseños de $N=1$ permiten una valoración objetiva y rigurosa del efecto de un tratamiento para un individuo—, mientras que la segunda cuestión —la eficacia del tratamiento para el paciente promedio— la responderemos mediante un diseño de grupos.

A estas situaciones cabe añadir dos más en las cuales un diseño de $N=1$ constituirá una mejor opción que un diseño de grupos:

1. Hay situaciones en las que no será conveniente recurrir a un diseño con grupo control por motivos éticos. ¿Es ético dejar de aplicar un tratamiento que se supone ventajoso a un grupo de pacientes con un trastorno psicológico por el hecho de que necesitemos un grupo control en nuestra investigación?
2. A veces podemos encontrarnos con problemas prácticos para reclutar un grupo amplio de pacientes que sean homogéneos respecto al trastorno conductual concreto que padecen.

2.3.2.2. Clasificación de los diseños de $N=1$

Antes de adentrarnos en la clasificación de los diseños de $N=1$, vamos a detenernos para exponer la notación que se usa habitualmente en éstos. Se han propuesto dos sistemas de notación. El primero lo plantearon Campbell y Standley (1966), y consiste en utilizar el símbolo O , para representar las observaciones de la variable dependiente, y el símbolo I , para representar que se ha introducido un tratamiento o intervención en un punto concreto de la serie. De este modo la representación del modelo básico del diseño experimental de $N=1$ es la siguiente:

$$O_1 O_2 O_3 I_1 O_4 O_5 O_6$$

A partir de esta representación básica se pueden introducir modificaciones, que pueden consultarse en el trabajo original de Campbell y Standley. El segundo sistema de notación sigue la tradición del análisis conductual aplicado (Barlow y

Hersen, 1984) y recurre a los siguientes símbolos: la letra A simboliza la línea base y las otras letras (B, C, D, etc.) simbolizan las fases en las que se introducen los distintos tratamientos. Este segundo sistema de notación es el que se utiliza con más frecuencia, por lo que lo seguiremos para presentar los distintos tipos de diseños.

Vamos a presentar la clasificación de los diseños de N=1 planteada por Hayes (1981) y Arnau (1994; 1995b), y que, recientemente, también han adoptado Ato y Vallejo (2007) y Balluerka (2001). No obstante, queremos destacar que se pueden encontrar otros sistemas de clasificación para este tipo de diseños (véase, por ejemplo, Kratochwill, 1992).

Los diseños de N=1 se clasifican en función de dos criterios. Primero: En función de la reversibilidad de la respuesta a los niveles de la línea base tras retirar o alterar el tratamiento. Segundo: En función de la estrategia de comparación que se utiliza para estudiar la efectividad del tratamiento. El primer criterio diferencia entre los *diseños de reversión* y los *diseños de no reversión*. Los diseños de reversión son aquellos en los que la conducta es susceptible de regresar al nivel de la línea base después de la intervención. Los diseños de no reversión, por el contrario, son aquellos en los que la conducta no puede volver al nivel de la línea base. El segundo criterio diferencia entre *diseños intra-serie*, *diseños entre-series* y *diseños de series mixtas o combinadas*. En los diseños intra-serie se registra la conducta en una sola serie de datos —por ejemplo, un solo individuo— y la efectividad de la intervención se valora mediante la comparación entre las sucesivas fases de esa serie. En los diseños entre-series se registra la conducta en dos o más series de datos y la efectividad de la intervención se valora mediante la comparación entre series de datos distintas. En los diseños mixtos se registra la conducta en dos o más series de datos y la efectividad de la intervención se valora mediante comparaciones intra-serie y entre-series.

Sin ánimo de ser exhaustivos vamos a presentar ahora algunos de los diseños de N=1 que se utilizan con más frecuencia. En la Tabla 2.7 se presentan clasificados en función de los dos criterios anteriores.

Tabla 2.7. Clasificación de los diseños experimentales de N=1 (adaptada de Ato y Vallejo, 2007).

Diseños reversibles	Diseños intra-series	Diseño de reversión A-B-A Diseño de reversión A-B-A con refuerzo diferencial de otras conductas Diseño A-B-A-B Diseño de retirada B-A-B Diseño de reversión multinivel A-B ₁ -A-B ₂ -A Diseño de tratamiento múltiple A-B-A-C-A Diseño de interacción A-B-A-B-BC-B-BC
	Diseños de series mixtas	Diseño de reversión de cuatro fases de inversión Diseño de reversión de cuatro fases de generalización
Diseños no reversibles	Diseños intra-series	Diseño de cambio de criterio
	Diseños entre-series	Diseño de tratamientos alternantes Diseño de tratamientos simultáneos
	Diseños de series mixtas	Diseño de línea base múltiple entre conductas Diseño de línea base múltiple entre participantes Diseño de línea base múltiple entre contextos

Antes de explicar los diseños de la Tabla 2.7, vamos a detenernos en el diseño A-B. Se trata del esquema más básico de un diseño de N=1. Consiste en el registro de la variable dependiente en una sola serie o unidad de análisis a lo largo de dos fases: una primera de línea base y otra de intervención. Estrictamente no se considera un diseño experimental, puesto que hay muchos factores que pueden atentar contra la validez interna de la investigación. El problema básico estriba en que si se observa un cambio al pasar de la línea base a la fase de tratamiento, este cambio puede deberse al impacto del tratamiento, pero también a otros factores, como, por ejemplo, la historia, la maduración, etcétera (una descripción de los distintos factores que pueden atentar contra la validez interna de una investigación se presentará más adelante en el epígrafe que corresponde a los diseños cuasi-experimentales). En definitiva, el diseño A-B no nos permitirá establecer una relación causal entre la administración del tratamiento y el cambio en la conducta.

Aunque el diseño A-B no es un diseño experimental, constituye la estructura básica a partir de la cual se construyen los diseños experimentales de N=1. Estos diseños son los que se exponen a continuación:

Diseños reversibles intra-serie: Se utiliza este tipo de diseños cuando se introduce una fase crítica o de reversión del tratamiento, y la efectividad del tratamiento se evalúa mediante la comparación de las distintas fases en una sola serie de datos.

- ♦ Diseño de reversión A-B-A. Se trata de un diseño con tres fases: línea base, tratamiento y reversión. Dada su estructura supera algunos de los problemas de validez interna que están presentes en el diseño A-B; no obstante, es un diseño que se utiliza poco en la práctica clínica porque acaba con una fase en la que se retira el tratamiento. Es evidente que si la unidad de análisis es un paciente, no es deseable que su conducta revierta a su estado inicial o línea base. Este tipo de diseño presenta otro inconveniente que es la posible presencia de efectos residuales. Estos efectos ya se han comentado cuando nos hemos referido a los diseños de medidas repetidas.
- ♦ Diseño de reversión A-B-A con refuerzo diferencial de otras conductas. Se trata de un diseño con tres fases: 1ª) línea base; 2ª) fase de tratamiento, en la que se refuerza la conducta de interés; y 3ª) fase crítica, en la que el tratamiento se retira de la conducta reforzada en la fase de tratamiento y se aplica a una conducta alternativa incompatible con la anterior. El principal inconveniente de este tipo de diseño es que, igual que en el caso anterior, acaba con una fase en la que no se refuerza la conducta de interés.
- ♦ Diseño A-B-A-B. Se conoce también con el nombre de diseño de retirada de cuatro fases o diseño A-B replicado. Sus cuatro fases son las siguientes: línea base, tratamiento, reversión y replicación del tratamiento. Por su estructura se trata de un diseño experimental en el que se incrementa considerablemente la validez interna si lo comparamos con los diseños anteriores. El efecto de la intervención estará claro si la conducta de interés cambia en el sentido esperado en la fase de tratamiento (primera B), vuelve a la línea base tras la retirada del tratamiento (segunda A) y se produce de nuevo la mejora en la última fase (segunda B). Además de su elevada validez

interna, este diseño presenta la ventaja de terminar en una fase de tratamiento, por lo que se utiliza con frecuencia en el ámbito clínico.

Una variante de este diseño es el diseño A-B-A₁-B, que se caracteriza porque en la tercera fase se retira el tratamiento B, pero se introduce un tratamiento placebo. Esta tercera fase permite controlar la reactividad del sujeto.

- ♦ Diseño de retirada B-A-B. Se utiliza cuando por consideraciones clínicas es necesario una intervención inmediata sin poder establecer una línea base. El diseño comienza con la aplicación de un tratamiento, que luego se retira y después se vuelve a aplicar. La retirada y posterior introducción del tratamiento permiten determinar si los cambios en la conducta son debidos al tratamiento. Este diseño también se utiliza cuando el nivel de la línea base es obvio porque la conducta objetivo no ha ocurrido nunca. Dado que la proporción de ocurrencia de la conducta objetivo es cero antes de la intervención no tiene sentido registrar una línea base.
- ♦ Diseño multinivel A-B₁-A-B₂-A y diseño de tratamiento múltiple A-B-A-C-A. Estos diseños resultan de la combinación de dos diseños A-B-A y se utilizan para comprobar, dentro de un mismo estudio, cómo dos o más tratamientos cuantitativa o cualitativamente diferentes afectan a la conducta de interés. En el diseño multinivel se combinan dos o más niveles de un mismo tratamiento, mientras que en el diseño de tratamiento múltiple se combinan dos o más tratamientos diferentes.
- ♦ Diseño de interacción. Es un diseño que, a diferencia de los anteriores, presenta una estructura compleja. Si bien existen muchas modalidades para representarlo, su orden básico se simboliza así: A-B-A-C-A-BC. Puede verse en esta estructura que este diseño consiste en la presentación de una serie de fases en las que cada tratamiento se administra solo y en combinación con otros. Siempre se comienza con la administración individual de los tratamientos y se continúa con la administración combinada de éstos. El objetivo de este diseño no es sólo evaluar el efecto aislado e independiente de cada variable de tratamiento, sino también determinar el efecto combinado o conjunto de éstas —es decir, determinar si la administración de un tratamiento influye en la efectividad de otro—. Sin embargo, el problema es que para estudiar la acción combinada de las variables se han de incluir muchas fases en el diseño (pensemos por ejemplo en el caso de trabajar con

tres variables distintas), lo cual hace que, en muchas ocasiones, sea imposible llevarlos a la práctica.

Diseños reversibles de series combinadas o mixtas: Se trata de diseños en los que se introduce una fase crítica o de reversión del tratamiento, se trabaja con dos o más series de datos y, además, la valoración de la eficacia del tratamiento se efectúa mediante comparaciones intra-serie y entre-series.

- ♦ Diseño de reversión de cuatro fases de inversión. También se conoce con el nombre de *diseño A-B-A-B con técnica de inversión*. Consiste en lo siguiente: 1º) durante la fase de línea base se registran simultáneamente dos conductas incompatibles de un mismo individuo: una que queremos incrementar y otra incompatible con ésta; 2º) se aplica el tratamiento a la conducta que se quiere incrementar; 3º) se invierte la aplicación del tratamiento, es decir, se aplica a la conducta incompatible con la primera; y 4º) se vuelve a aplicar el tratamiento a la conducta que se quiere incrementar. En este caso se realiza la comparación entre las fases de cada serie por separado (comparación intra-serie) y la comparación entre las mismas fases para cada conducta (comparación entre-series).
- ♦ Diseño de reversión de cuatro fases de generalización. Este diseño también se conoce como *diseño A-B-A-B con registro de conductas concurrentes* (Barlow y Hersen, 1984). Consiste en el registro simultáneo de la conducta que se quiere modificar y de otras conductas concurrentes o covariantes a la misma. Si todas las conductas son equivalentes cabe esperar que el tratamiento afecte a todas ellas. Si, por el contrario, todas las conductas no son equivalente, entonces se espera que el tratamiento afecte a la conducta de interés y a las relacionadas con ésta, pero no al resto de conductas. La ventaja de este tipo de diseño es que nos permite estudiar el impacto del tratamiento en otras conductas distintas de la conducta de interés más inmediato.

Diseños no reversibles intra-serie: Estos diseños se utilizan cuando no es posible revertir el efecto del tratamiento, por razones técnicas (efectos residuales o irreversibles de la intervención) o por consideraciones éticas. Además, se trata de

diseños en los que se utiliza una sola serie de datos para valorar la eficacia de la intervención.

- ♦ Diseño de cambio de criterio. Este diseño comienza con una fase de línea base en la que se registra la conducta que se quiere modificar. A partir de aquí se van introduciendo fases sucesivas en las que el tratamiento se va administrando en función del nivel de ejecución del sujeto en la fase anterior. Este nivel de ejecución en la fase previa se conoce como criterio, y va cambiando de forma gradual entre las fases. En efecto, una vez se alcanza el criterio para una fase, se administra de nuevo el tratamiento y se establece un nuevo criterio. El nuevo criterio será superior o inferior al criterio anterior, dependiendo de si nuestro interés se centra en incrementar o reducir la conducta objetivo. La efectividad del tratamiento quedará demostrada en la medida en que la conducta se vaya ajustando a los distintos cambios de criterio y finalmente alcance el nivel deseado (que puede ser la eliminación completa de esa conducta). Las ventajas de este diseño son muchas —fácil aplicación, plasticidad, no se deja al paciente en una fase de reversión, etc.—. Sin embargo, presenta el inconveniente de necesitar una relativa estabilidad entre las fases para poder valorar la efectividad del tratamiento sin ambigüedad.

Diseños no reversibles entre-series: Se trata de diseños en los que no hay fase de reversión, y en los que para valorar la eficacia del tratamiento se compara la conducta entre series distintas.

- ♦ Diseño de tratamientos alternantes. Este diseño ha recibido otras denominaciones como diseño de programa múltiple (Barlow y Hersen, 1984), diseño de línea de base de elementos múltiples (Sidman, 1960; Ulman y Sulzer-Azaroff, 1975) y diseño de aleatorización (Edgington, 1967; 1972). Se utiliza para valorar la efectividad de dos o más tratamientos que se administran a un mismo individuo. Estos tratamientos se administran al individuo de forma alterna y rápida. La lógica de la estructura de este diseño es que una misma conducta presenta cursos relativamente independientes cuando se trata diferencialmente en situaciones distintas o en presencia de diferentes estímulos físicos y sociales (Leitenberg, 1973). El objetivo que se

persigue con este diseño no es sólo examinar si los tratamientos influyen en la conducta del individuo, sino que también se pretende descubrir cuál de los tratamientos es más efectivo.

- ♦ Diseño de tratamientos simultáneos. En este diseño, a diferencia del anterior, los tratamientos se presentan de forma simultánea o concurrente al sujeto (Kazdin y Hartman, 1978). Se trata de un diseño que se utiliza con poca frecuencia.

Diseños no reversibles de series combinadas o mixtas: Se trata de diseños en los que el tratamiento no se retira, se registran múltiples series y se evalúa la eficacia de la intervención mediante comparaciones intra y entre-series.

- ♦ Diseño de línea base múltiple. La característica fundamental de estos diseños es que la variable de tratamiento se aplica de forma escalonada en el tiempo a distintas conductas de un mismo individuo (diseño de línea base múltiple entre conductas), a la misma conducta para un individuo en situaciones diferentes (diseño de línea base múltiple entre contextos) o a la misma conducta en individuos diferentes (diseño de línea base múltiple entre sujetos). En consecuencia, se utilizarán dos o más líneas base que pueden ser conductas distintas, situaciones distintas o individuos distintos. Para cada línea base, el tratamiento se introducirá en un momento temporal diferente. Cualquiera que sea la versión del diseño de línea base múltiple, la forma de proceder es la siguiente: 1º) La primera intervención se introducirá cuando todas las líneas bases sean estables; entonces, se introducirá el tratamiento a la primera línea base, esperando que se produzca un cambio en la conducta de esta primera línea pero que no se produzca cambio alguno en las otras; 2º) Una vez que se hayan vuelto a estabilizar todas las series, entonces se introducirá el tratamiento en la segunda línea base, esperando que se produzca un cambio en la misma pero que no se produzca ninguna modificación en la conducta de las otras series; 3º) El mismo proceso se va repitiendo hasta alcanzar a la última línea base. La eficacia del tratamiento se demostrará si la conducta cambia en cada serie sólo a partir del momento en que se introduce el tratamiento.

2.3.2.3. Análisis de datos para este tipo de diseños

El análisis de datos para los diseños de $N=1$ se suele realizar de dos maneras: (1) mediante la inspección visual de los datos y (2) a través del análisis estadístico. Dada la vinculación de este tipo de diseño con la práctica clínica y educativa, el primer tipo de análisis es el que se ha aplicado con más frecuencia, hecho que ha fomentado el desarrollo de una serie de técnicas y procedimientos para el análisis visual (véase una revisión de éstas en Franklin, Gorman, Beasley y Allison, 1996). Sin embargo, los análisis visuales también han recibido muchas críticas, especialmente por su dudosa fiabilidad (Knapp, 1983; Wampold y Furlong, 1981). Tales críticas han favorecido la propuesta del uso de técnicas estadísticas como complementarias e incluso sustitutivas de este tipo de análisis. Aunque al principio algunos autores propusieron el uso de pruebas estadísticas clásicas para el análisis de los datos de este tipo de diseño (Gentile, Roden y Klein, 1972; Shine y Bower, 1971), pronto se descartó su utilización por los problemas que esto suponía — básicamente el incumplimiento de la condición de independencia de los términos de error debido a la dependencia serial de los datos—. Esta es la razón por la que se recurrió a otras técnicas estadísticas: unas de tipo paramétrico, como los modelos ARIMA (Box y Jenkins, 1970; Box y Tiao, 1965), y otras de tipo no paramétrico, basadas en los principios de la aleatorización (Edgington, 1995). Sin embargo, estas técnicas estadísticas no siempre constituyen una buena alternativa a las técnicas basadas en la inspección visual para el análisis de este tipo de datos. Balluerka (2001) destaca sus dos principales inconvenientes. El primero es que estas técnicas estadísticas requieren el cumplimiento de una serie de condiciones que resultan difíciles de satisfacer en los contextos aplicados; el segundo es que en muchas ocasiones el analista no busca la significación estadística, sino la significación clínica —por ejemplo, que se produzca una mejora importante en el funcionamiento cotidiano del paciente—. Estas son las principales razones por las que la utilización de las técnicas estadísticas no se ha extendido en el análisis de datos para diseños de $N=1$. En definitiva, por el momento, no existe un procedimiento ideal de análisis para evaluar el impacto del tratamiento en este tipo de diseños. La sugerencia de la mayoría de los autores (véase, por ejemplo, Arnau, 1995d y Balluerka, 2001) es que se recurra tanto al análisis visual como al análisis estadístico para alcanzar una evaluación más completa de los resultados.

Puesto que una exposición detallada de todas las técnicas de análisis excede los objetivos de este texto, vamos a centrarnos en las técnicas del análisis visual. Nos detendremos en éstas por tres motivos. En primer lugar, porque a pesar de las críticas que ha recibido el análisis visual es el que más se utiliza en el ámbito aplicado (Arnau, 1995d; Kratochwill y Brody, 1978; Poling *et al.*, 1995; Wampold y Furlong, 1981). En segundo lugar, porque se ha demostrado que el entrenamiento en el análisis visual de datos mejora la interpretación de éstos. Fisher, Kelley y Lomas (2003) informaron de que un grupo de observadores no entrenados interpretaba los gráficos con una precisión de aproximadamente el 55% respecto al criterio de un grupo de expertos, y que esos mismos observadores incrementaban su ejecución hasta aproximadamente el 95 % después de recibir entrenamiento en la inspección visual. En tercer lugar, nos vamos a centrar en el análisis visual porque las técnicas de análisis estadístico para este tipo de datos constituyen una materia ajena a la que ahora nos ocupa. Por lo que se refiere a estas técnicas estadísticas pueden encontrarse en Arnau (1995d, 2001), Ato y Vallejo (2007), y Gorman y Allison (1996), entre otros.

Vamos a abordar el tema del análisis visual de los datos deteniéndonos en dos aspectos esenciales. Primero: expondremos de qué manera se ha de construir un gráfico para que pueda interpretarse sin ambigüedad. Segundo: indicaremos qué elementos del gráfico han de captar la atención del investigador y cómo deberá descifrarlos para poder alcanzar una adecuada interpretación de los resultados.

Para construir un gráfico con los resultados de un diseño de N=1 se suele recurrir a un diagrama o gráfico de curvas. El eje de abscisas se reserva para los periodos de observación (ocasiones, sesiones o tiempo) y el eje de ordenadas se reserva para la variable dependiente o conducta. Por lo que se refiere a la disposición en el gráfico de otras informaciones relevantes vamos a seguir las directrices para la elaboración de gráficos formuladas por la Sociedad para el Análisis Experimental de la Conducta, que se publican periódicamente en el *Journal of Applied Behavior Analysis* (pueden encontrarse también en Franklin, Gorman, Beasley y Allison, 1996).

1. Las variables independientes se identificarán mediante rótulos o etiquetas colocados encima de los datos.
2. Los niveles o condiciones de la variable independiente se separarán mediante líneas discontinuas perpendiculares al eje de abscisas.

3. Los puntos del gráfico se unirán mediante líneas rectas para facilitar la visualización de la variabilidad y las tendencias presentes en los datos. Los puntos se han de destacar más que las rectas para que sean claramente visibles.
4. Cuando se muestren dos o más series de datos en un mismo gráfico —que representen, por ejemplo, la misma conducta para sujetos distintos o conductas distintas para un mismo sujeto— éstas deberán aparecer convenientemente etiquetadas.
5. El nivel de respuesta cero no deberá descansar sobre el eje de abscisas, sino que deberá levantarse ligeramente.
6. La escala de medida de la variable dependiente deberá ser adecuada para el tipo de respuesta registrada a fin de evitar las interpretaciones incorrectas debidas a los cambios de escala.
7. Si la escala en el eje de ordenadas no es continua se ha indicar mediante la presentación de una rotura o interrupción en el eje. Esta directriz es controvertida porque las interrupciones en el eje de ordenadas pueden provocar la sensación de que las diferencias entre las respuestas son mayores de lo que realmente son.

Una vez que el investigador ha construido su gráfico deberá proceder con el análisis visual de sus datos. Para ello evaluará tanto el patrón que presentan sus datos intra-fase como el patrón que presenten sus datos entre-fases (Kennedy, 2005). Para juzgar el impacto de la intervención, el investigador podrá recurrir a algunas técnicas de estadística descriptiva —por ejemplo, al análisis exploratorio de datos propuesto por Tukey (1977)—. Estas técnicas estadísticas, a diferencia de las que hemos mencionado anteriormente, son descriptivas, por lo que no nos permitirán hacer inferencias poblacionales. Vamos a presentar ahora los elementos básicos a los que se ha de atender para evaluar el patrón intra-fase y el patrón entre-fases, así como algunas de las técnicas de estadística descriptiva que pueden resultar útiles para resumir o sintetizar la información.

Patrones intra-fase: Hay tres elementos que se han de considerar cuando se está valorando un patrón intra-fase mediante la inspección visual.

El primer elemento es el *nivel de los datos*. Un indicador estadístico de este nivel puede ser cualquier medida de tendencia central que se calcule con los datos de la

fase. Aunque se puede recurrir a la media aritmética, su uso no es muy recomendable, puesto que en muchas ocasiones se trabaja con series cortas —que incluyen pocos datos—, y éste es un índice muy sensible ante la presencia de datos anómalos —en inglés, *outliers*—. Esta es la razón por la que es aconsejable recurrir a medidas de tendencia central que sean resistentes, como por ejemplo, la mediana, las medias recortadas o los resúmenes promedio.

El segundo elemento a considerar cuando se procede con un análisis visual es la *tendencia de los datos*. Esta tendencia puede definirse como la línea recta que se ajusta mejor a los datos de una fase. En este momento, el investigador deberá dibujar esa línea en su gráfico y juzgar la inclinación y la magnitud de ésta. Por una parte, la inclinación se valorará como positiva —si los valores de la variable de respuesta incrementan su valor dentro de la fase (la recta es ascendente)—, negativa —si los valores de la variable de respuesta decrecen dentro de la fase (la recta es descendente)— y plana —si los valores de la variable de respuesta se mantienen constantes dentro de la fase (la recta es una línea horizontal)—. Por otra parte, la magnitud se valorará en función del tamaño de la pendiente de la recta. Así, por ejemplo, se dirá que la magnitud es baja si los valores de la variable de respuesta crecen o decrecen lentamente. Dado que la tendencia se valora tras ajustar una línea recta a los datos de la fase, en este momento se podrá recurrir a algunas técnicas estadísticas que nos permitirán obtener la recta que mejor se ajuste a los datos. Franklin *et al.* (1996) presentan tres técnicas que se pueden utilizar: la regresión lineal por mínimos cuadrados, el método del *split-middle* y el ajuste de línea de tendencia resistente. En este momento hay que aclarar que el ajuste de una línea recta a los datos no siempre será la mejor opción, puesto que, en ocasiones, podemos encontrar tendencias no lineales en los datos. Estas tendencias no lineales pueden ser debidas, por ejemplo, a cambios cíclicos en la conducta de interés.

El tercer elemento que se utiliza para juzgar el patrón intra-fase es la *variabilidad*. Esta variabilidad nos informa de la dispersión de los datos; en otras palabras, es una medida de la homogeneidad o heterogeneidad de éstos. Como norma general, interesa que los datos presenten poca variabilidad intra-fase, puesto que esto es un indicador de que el patrón de datos es estable. Morley y Adams (1991) proponen dos métodos para visualizar la información acerca de la variabilidad: las *barras de rango* (Parsonson y Baer, 1978) y las *líneas de rango*. Las barras de rango se construyen conectando una línea vertical a través de tres puntos

dentro de cada fase: una estimación de la tendencia central, el valor máximo y el valor mínimo. Las líneas de rango se construyen dibujando en el gráfico dos líneas horizontales, cada una de ellas para los extremos del rango. Estos dos métodos se han criticado por tres razones: (1) son muy sensibles a la presencia de datos extremos; (2) la variabilidad se puede confundir con la tendencia; y (3) no recogen los cambios en la tendencia dentro de las fases.

Una mención especial merece la evaluación del patrón intra-fase de la línea base. Kazdin (2001) destaca que la línea base tiene una doble función. De una parte, proporciona una descripción del nivel existente de la conducta. De otra parte, permite predecir la conducta para el futuro inmediato en caso de que no se introduzca ningún tratamiento. Esta función predictiva pone a nuestra disposición un estándar a partir del cual se puede evaluar la eficacia consiguiente de la intervención. Como la línea base es el elemento básico de comparación con la fase de tratamiento, es esencial que presente un patrón estable y libre de tendencias — como mucho se puede aceptar que la línea base presente una tendencia contraria a la que se espera que se produzca tras la intervención—. Si el patrón de la línea base no cumple estos requisitos, entonces es muy difícil, si no imposible, determinar si se produce algún cambio en la fase de tratamiento. Efectivamente, en este caso sólo se podrá concluir que el tratamiento es eficaz si el impacto del tratamiento sobre la conducta es suficientemente contundente como para que se detecte. En general, se recomienda que las medidas de la línea base se prolonguen en el tiempo hasta conseguir que se estabilice o hasta que presenten una tendencia contraria a la esperada tras la intervención (Baer, Wolf y Risley, 1968; McNamara y MacDonough, 1972). No obstante, Barlow y Hersen (1984) han informado de los problemas que puede acarrear el incremento de las observaciones en la línea base; entre otros, apuntan la consideración ética que supone demorar la administración de un tratamiento que se supone que mejoraría los síntomas —o incluso sanaría completamente— a un paciente.

Patrones entre-fases: La comparación de los patrones entre-fases permitirá al investigador juzgar el efecto de su intervención. Para esto, además de valorar si se producen cambios de nivel o de tendencia entre las fases, puede atender a otros dos elementos informativos. Primero: la *inmediatez del efecto* de la intervención (o rapidez del cambio). Segundo: el *solapamiento* entre fases.

La inmediatez del efecto de la intervención se refiere a si el cambio —de nivel o de tendencia— en la variable de respuesta que se observa tras la intervención ocurre inmediatamente después de ésta o si, por el contrario, este cambio se demora. En general, cuanto menos tiempo transcurre entre la introducción del tratamiento y el cambio, más convincente será la relación funcional entre ambos.

El solapamiento puede definirse como el porcentaje de valores de la variable de respuesta que comparten las dos fases que se quieren comparar. En otras palabras, *el solapamiento se produce cuando los puntos de una fase (por ejemplo, la fase de intervención) son asumibles por los de la fase anterior* (Arnau, 1995d, p. 198). En general, la eficacia de la intervención será más clara en la medida en que no se produzca solapamiento entre la fase de línea base y la fase de tratamiento. No obstante, hay un caso en el que aunque ambas fases no se solapen esta efectividad no estará clara; esto se producirá cuando haya una tendencia creciente o decreciente continua entre las dos fases.

Las recomendaciones que se han dado anteriormente para llevar a cabo la inspección visual de los datos de un diseño de $N=1$ serán, en general, apropiadas para valorar el impacto de un tratamiento siempre que la variabilidad dentro de las fases sea pequeña. No obstante, como señalan Ato y Vallejo (2007) *la ambigüedad de las inferencias visuales irá en aumento conforme lo haga la variabilidad dentro de las fases, las magnitudes del efecto decrezcan y los tratamientos evaluados sean más novedosos* (p. 575). En esos casos, los análisis visuales necesariamente deberán complementarse con otros análisis estadísticos.

2.3.3. Diseños cuasi-experimentales

2.3.3.1. Definición y características

El concepto de cuasi-experimento fue propuesto por primera vez por Campbell y Stanley (1966) y fue ampliado más tarde por Cook y Campbell (1979). A partir de entonces muchos autores han propuesto definiciones de este concepto, algunas de las cuales figuran a continuación.

Cook y Campbell (1986) afirman que 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). Esta es la razón por la cual estos autores utilizan el término *experimento verdadero* en oposición al término *cuasi-experimento*.

Kirk (1995) afirma que los diseños cuasi-experimentales *son similares a los experimentos excepto en que los sujetos no se asignan aleatoriamente a la variable independiente. Se trata de diseños que se utilizan cuando la asignación aleatoria no es posible o cuando por razones prácticas o éticas se recurre al uso de grupos naturales o preexistentes como, por ejemplo, sujetos con una determinada enfermedad o sujetos que han sido sometidos a abuso sexual (p. 6).* Por lo tanto, los diseños cuasi-experimentales se utilizan cuando el investigador no puede presentar los niveles de la variable independiente a voluntad ni puede crear los grupos experimentales mediante la aleatorización.

Pedhazur y Pedhazur Schmelkin (1991) definen el cuasi-experimento como *una investigación que posee todos los elementos de un experimento, excepto que los sujetos no se asignan aleatoriamente a los grupos. Esta ausencia de aleatorización hace que el investigador se enfrente a la tarea de identificar y separar los efectos de los tratamientos de los efectos de otros factores que afectan a la variable dependiente (p. 277).*

Ato (1995a) afirma que *la metodología cuasi-experimental incorpora 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. 46).*

Arnau define el diseño cuasi-experimental como *un plan de trabajo con el que se pretende estudiar el impacto de los tratamientos y/o los procesos de cambio, en situaciones donde los sujetos o unidades de observación no han sido asignados de acuerdo con un criterio aleatorio.*

A partir de las definiciones anteriores se puede elaborar un listado con las principales características del diseño cuasi-experimental. Son las siguientes:

1. Manipulación de la variable independiente. Esta es una característica que comparten los diseños cuasi-experimentales y los diseños experimentales. Ambos tipos de diseño tienen como objetivo el estudio del efecto de la variable independiente sobre la variable dependiente de la investigación. En definitiva, los dos tipos de diseños persiguen el establecimiento de relaciones causales.

2. No aleatorización en la formación de los grupos. En el diseño cuasi-experimental el investigador no interviene en la formación de los grupos, de manera que recurre a grupos intactos o *naturales*. Se trata de grupos de individuos que ya están formados —como, por ejemplo, los niños de un mismo grupo en una escuela o los trabajadores de un departamento en una empresa—, por lo que el investigador no tiene garantías de la equivalencia inicial de éstos. De hecho, estos grupos naturales, también se denominan *grupos no equivalentes*. Esta característica constituye el principal inconveniente que presenta este tipo de diseños. El problema radica en que si la equivalencia inicial de los grupos no está garantizada, se puede cuestionar que las diferencias que se encuentren entre los grupos después de la intervención se deban al efecto del tratamiento. En definitiva, esta característica atenta contra la validez interna de la investigación, es decir, pone en cuestión la relación causal que se pretende establecer con ésta. Algunos metodólogos se refieren a este tipo de diseños como *estudios no aleatorizados* (Anderson *et al.*, 1980).
3. Escaso control de las variables de confundido. Los diseños cuasi-experimentales se suelen utilizar en investigaciones de carácter aplicado, por lo que se desarrollarán, principalmente, en contextos naturales alejados del laboratorio. En estos contextos el control de todas las variables de posible confundido resulta complicado y, en muchas ocasiones, imposible. Estos contextos naturales pueden ser, entre otros, hospitales, escuelas o empresas. Esta tercera característica, al igual que la anterior, pone en peligro la validez interna de la investigación. Debido al deficiente control de los diseños cuasi-experimentales, algunos autores, como Van Dalen y Meyer (1971), los denominan *investigaciones con control parcial*.

Las características anteriores ponen de manifiesto cuál es el principal problema de los diseños cuasi-experimentales respecto al diseño experimental. Si bien en el diseño experimental se alcanza un alto grado de validez interna, en el diseño cuasi-experimental, por el contrario, la validez interna de la investigación se puede cuestionar. Esto se debe a que los cuasi-experimentos no permiten elaborar conclusiones consistentes acerca de la efectividad del tratamiento, puesto que existen explicaciones alternativas, distintas del efecto de la intervención, para justificar las diferencias que se observan en la variable dependiente. En definitiva,

los diseños cuasi-experimentales no nos permiten rechazar completamente otras explicaciones causales distintas a la que nos interesa establecer. Esta es la razón por la que Campbell y Standley (1963, 1966) recomiendan la utilización de los diseños cuasi-experimentales únicamente cuando no se puedan asignar aleatoriamente los sujetos a las diferentes condiciones.

Según Kenny (1975, 1979) la no aleatorización en la formación de los grupos puede darse en las siguientes circunstancias: (1) el tratamiento se administra a un grupo —aula, colegio, sistema escolar— y se toma otro grupo como control; (2) se había planificado realizar un experimento verdadero, pero debido a la mortalidad o a la 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; (3) por escasez de recursos sólo se aplica el tratamiento a un grupo pre-seleccionado; y (4) los sujetos autoseleccionan su propio nivel de tratamiento.

Campbell y Standley (1966), y posteriormente Cook y Campbell (1979), señalaron cuáles son las potenciales amenazas contra la validez interna que pueden darse en los diseños cuasi-experimentales (también pueden encontrarse en Shadish, Cook y Campbell, 2002). Tales amenazas son las que presentamos a continuación:

1. Historia. Son hechos o circunstancias externas que ocurren simultáneamente con la aplicación del tratamiento. El problema es que en este caso el investigador no podrá estar seguro de que los cambios observados en la variable dependiente de su investigación se deban al tratamiento, puesto que podrían deberse al factor de historia.
2. Maduración. Se refiere a cambios en las condiciones internas de los individuos que coinciden con la aplicación del tratamiento. Se trata de cambios debidos a procesos biológicos o psicológicos, como, por ejemplo, la edad, la fatiga o el aburrimiento.
3. Efecto de la administración de pruebas previas. Este efecto se produce en los diseños en los que toman medidas pretest. En principio las diferencias que se observen entre las medidas pretest y las medidas posttest se tendrían que atribuir al efecto del tratamiento, no obstante, en ocasiones, una buena parte de esos cambios pueden deberse a la práctica o entrenamiento en el pretest.

4. Instrumentación. Se refiere a cambios que se producen en el calibrado de los instrumentos de medida o a problemas con los observadores que coinciden con la aplicación del tratamiento.
5. Regresión estadística. Se refiere a la tendencia hacia la centralidad que se produce cuando el criterio para formar los grupos es que los individuos presenten puntuaciones extremas. Dicho de otro modo, las puntuaciones extremas tienden a acercarse a la media en el posttest.
6. Selección diferencial de los sujetos. Se refiere al sesgo en la formación de los grupos y se produce cuando en la investigación se recurre a grupos naturales o intactos. El problema que se presenta en estos casos es que no queda garantizada la equivalencia inicial de los grupos, por lo que las diferencias que se observen entre éstos después de la intervención pueden deberse a la ausencia de equivalencia inicial y no al efecto del tratamiento.
7. Mortalidad selectiva. Se refiere a la pérdida no aleatoria de los individuos que forman parte de los grupos. Este problema aparece en los diseños en los que se requiere tomar dos o más medidas de cada individuo en momentos temporales diferentes. Puede ocurrir, en estos casos, que a medida que se desarrolla la investigación se vaya reduciendo el tamaño de la muestra, porque haya individuos para los que no se consiga tomar medidas de todas las observaciones previstas inicialmente.
8. Interacciones entre la selección y algunas de las amenazas anteriores. Las interacciones más frecuentes se producen entre la selección y la historia —por ejemplo, puede ocurrir que un factor de historia afecte sólo a uno de los grupos de la investigación, introduciendo un sesgo sistemático en la variable dependiente—, y entre la selección y la maduración —que se produce cuando los grupos maduran con un ritmo diferente—.
9. Ambigüedad acerca de la dirección de la inferencia causal. En algunas investigaciones puede ser difícil determinar si X es responsable del cambio en Y o viceversa. Esta ambigüedad no se producirá si sabemos que X ocurrió antes que Y.

Aunque reconocemos que la lectura del listado anterior puede desanimar a un investigador respecto a la utilización de los diseños cuasi-experimentales, queremos recordar que estas amenazas son *potenciales*, lo cual quiere decir que no

necesariamente se producirán en todo diseño cuasi-experimental. Tal como señalan León y Montero (1997) *no todas las amenazas a la validez interna tienen por qué presentarse siempre o, al menos, no todas a la vez* (p. 291). Por otra parte, si bien un diseño experimental superaría la mayoría de estas amenazas, también sabemos que no siempre es viable llevar a cabo una investigación de este tipo. En consecuencia, siempre será mejor obtener información acerca de un fenómeno, aunque no se disponga de una garantía total de la validez interna, que renunciar al estudio de ese fenómeno. En definitiva, lo que sí es esencial es que el investigador conozca las limitaciones del diseño que ha utilizado para recoger sus datos y actúe en consecuencia cuando se disponga a elaborar sus conclusiones.

2.3.3.2. Clasificación de los diseños cuasi-experimentales

Si bien existen distintas clasificaciones de los diseños cuasi-experimentales, la mayoría de ellas coinciden en proponer dos grupos de diseños. De una parte, están los diseños transversales, y, de otra parte, están los diseños longitudinales (veáanse por ejemplo: Arnau, 1997, 2003; Ato y Vallejo, 2007; Judd y Kenny, 1981). En concreto, vamos a presentar aquí la clasificación propuesta por Arnau (1997, 2003). Los diseños transversales se caracterizan porque estudian el fenómeno en un momento temporal concreto, de manera que la variable de respuesta o variable dependiente se mide en un único momento temporal —como mucho tomaríamos otra medida de la variable de respuesta antes de la intervención o, lo que es lo mismo, tomaríamos una medida pretest o pretratamiento—. Se trata de diseños en los que se comparan grupos. Los diseños longitudinales, por el contrario, se caracterizan porque en ellos se toman varias medidas de la variable de respuesta para los distintos individuos —que pueden ser uno solo o más de uno (aulas, escuelas, poblaciones)— a lo largo del tiempo. Su objetivo es estudiar los procesos de cambio en función del tiempo y explicarlos.

Los diseños transversales se clasifican a su vez en función de la regla de asignación de los sujetos a los grupos. Así, se distingue entre los diseños con regla de asignación desconocida —el *diseño de grupo control no equivalente* y el *diseño de grupos no equivalentes*— y los diseños con regla de asignación conocida —el *diseño de discontinuidad en la regresión*—. Los diseños longitudinales se clasifican en *diseños de series temporales interrumpidas*, *diseños de medidas repetidas*,

diseños de cohortes y diseños en panel. Vamos a presentar ahora una breve descripción de cada uno de estos diseños.

El diseño de grupo control no equivalente, en su forma básica, se caracteriza porque utiliza dos grupos: uno recibe el tratamiento, intervención o programa cuyo efecto se pretende estudiar, y otro no recibe tratamiento o recibe un tratamiento placebo (Ato y Vallejo, 2007). Así se dispone de un grupo experimental y un grupo control. En su forma más simple se toman medidas sólo después de la intervención —*el diseño de grupo control no equivalente con medidas sólo posttest*—. No obstante, en este caso, los problemas de control son tan grandes que muchos autores no incluyen este diseño en la categoría de cuasi-experimento, sino en la de pre-experimento. Si al diseño anterior se le añaden unas medidas pretest, alcanzará la categoría de diseño cuasi-experimental. En este caso, las medidas pretest permitirán, en primer lugar, valorar la equivalencia inicial de los grupos, y, en segundo, controlarla en caso de que se dé la no equivalencia. Este diseño se denomina *diseño de grupo control no equivalente con pretest y posttest*.

El diseño de grupos no equivalentes es similar al diseño anterior, no obstante, en este caso, los grupos reciben tratamientos distintos. Además, se trata de grupos naturales, por lo que se tomarán medidas pretest como elemento de control de la equivalencia de los mismos.

El diseño de discontinuidad en la regresión se caracteriza porque, a diferencia de los anteriores, se conoce la variable de asignación de los sujetos a los grupos (Ato, 1995b; Trochim, 1984, 2006). Esta variable de asignación es una medida pretest. La forma de proceder para formar los grupos es la siguiente: una vez tenemos la medida pretest para todos los sujetos, éstos se ordenan en función de esa medida y se establece una puntuación de corte para asignar los sujetos a los grupos. Los sujetos con una puntuación por encima del punto de corte formarán parte de uno de los grupos y los sujetos con una puntuación por debajo del punto de corte pasarán a formar parte del otro grupo. Se trata de un diseño que goza de mucho prestigio porque incrementa las posibilidades interpretativas respecto a los otros diseños cuasi-experimentales, aproximándose incluso al diseño experimental (Shadish, Cook y Campbell, 2002). Muchos autores consideran que este diseño está a medio camino entre el diseño experimental clásico y el diseño cuasi-experimental típico (Ato, 1995c). Sin embargo, a pesar de sus ventajas, es un tipo de diseño que se ha utilizado poco en la práctica.

Los diseños de series temporales interrumpidas presentan como característica fundamental el registro de múltiples medidas previas y múltiples medidas posteriores a la introducción de un tratamiento o de un programa de intervención. Dentro de este grupo de diseños se distingue, a su vez, entre el *diseño de series temporales simple* y el *diseño de series temporales con grupo control no equivalente* (Arnau, 1995e). El primer diseño es aquél en el que se trabaja con un solo grupo de individuos, mientras que el segundo es aquél en el que se emplea un grupo experimental y un grupo control, con la finalidad de valorar de forma más rigurosa si los cambios producidos entre las fases pre y postintervención son iguales en ambos grupos. Este segundo diseño tiene mayor validez interna que el primero.

Los diseños longitudinales de medidas repetidas son aquellos en los que se registran más de dos medidas repetidas en el tiempo con el propósito de conocer el proceso de crecimiento de una muestra de sujetos. En estos diseños lo que se hace es modelar los datos con ajustes de polinomios con el fin de estudiar las curvas de crecimiento. Este tipo de diseño también se denomina *diseño longitudinal de múltiples observaciones* (1GMO). Un tipo especial de diseño longitudinal de medidas repetidas es el *diseño split-plot* o de muestra dividida. En este diseño se trabaja con dos o más grupos de sujetos que se han formado en función de una variable de carácter social, biológica o psicológica, y se toman medidas para cada individuo en momentos temporales diferentes.

El diseño longitudinal de cohortes nos permite estimar los efectos de la edad, el periodo y la cohorte en un proceso de cambio. Una cohorte es un grupo o agregado de individuos, dentro de una población específica, que ha experimentado el mismo acontecimiento vital en un intervalo de tiempo determinado (Gleen, 1977; Ryder, 1965; Visser, 1985). Aunque la variable más usada para delimitar la cohorte es la fecha de nacimiento, pueden utilizarse otras: los individuos casados en un período de tiempo establecido, los individuos que asistieron a la escuela en un determinado año, etc.

El diseño longitudinal en panel consiste, en la situación más simple, en medir dos variables de una misma muestra de sujetos a lo largo de una serie de tandas o momentos históricos. El objetivo es establecer, mediante la técnica de correlación cruzada en panel, la posible relación causal entre las variables medidas y el sentido de la causalidad. Se trata de un tipo de diseño que se utiliza mucho en el ámbito social para el estudio de intenciones políticas, preferencias, actitudes u opiniones.

2.3.3.3. Análisis de datos para este tipo de diseños

Las técnicas de análisis que se utilizan con más frecuencia en el análisis de datos de diseños cuasi-experimentales son el análisis de la variancia para datos independientes (AVAR), el análisis de la variancia para medidas repetidas (AVAR-MR), el análisis de la covariancia (ACOVAR), el análisis de la regresión múltiple (ARM), el análisis de series temporales (modelos ARIMA), el análisis de mínimos cuadrados generalizados (MCG), el análisis multivariado de la variancia (AMVAR), el análisis de las correlaciones cruzadas en panel (ACCP) y los modelos lineales de ecuaciones estructurales (LISREL). En la Tabla 2.8 se relacionan los distintos tipos de cuasi-experimentos con sus correspondientes técnicas de análisis.

Tabla 2.8. Diseños cuasi-experimentales y sus correspondientes técnicas de análisis (Bono, 1997).

Diseños	Técnicas de análisis
Diseño de grupo control no equivalente	AVAR, ACOVAR, ARM
Diseño de grupos no equivalentes	AVAR, ACOVAR, ARM
Diseño de grupos no equivalentes con múltiples covariables	AVAR, ARM, ACOVAR MÚLTIPLE
Diseño de discontinuidad en la regresión	AVAR, ACOVAR, ARM
Diseño de series temporales interrumpidas	ARIMA, MCG, LISREL
Diseño de medidas repetidas	AVAR-MR, AMVAR
Diseño de cohortes	AVAR, AVAR-MR
Diseños en panel	ACCP, AR, LISREL

Una exposición detallada de todas estas técnicas de análisis excedería los objetivos de este texto (véanse exposiciones detallada de éstas, por ejemplo, en: Arnau, 2001; Ato, 1995c, 1995d; Vallejo, 1995a, 1995b, 1996), por lo que presentaremos aquí sólo una de ellas: el análisis de la covariancia. Hemos escogido esta técnica porque se trata de una de las más utilizadas en los diseños cuasi-

experimentales de tipo transversal, en los que se recurre a ella como técnica de control estadístico de las diferencias iniciales entre los grupos. El objetivo de esta técnica estadística es el mismo que el de un AVAR, es decir, la estimación de los efectos del tratamiento; no obstante, en el ACOVAR esta estimación se realiza una vez que se ha ajustado el efecto de la covariable.

Si bien Fisher (1932) desarrolló el ACOVAR como método para incrementar la precisión de las estimaciones de los efectos de los tratamientos en experimentos aleatorizados, más tarde esta técnica se comenzó a utilizar como método para ajustar las diferencias iniciales entre los grupos en cuasiexperimentos (Maxwell, O’Callaghan y Delaney, 1993). En este segundo caso se utilizan las medidas preintervención —o pretest— como covariable, y la cuestión básica que se pretende responder es: ¿Los grupos serían diferentes en el postest si hubieran sido equivalentes en la covariable? (Maxwell y Delaney, 1990). En definitiva, se trata de eliminar la variabilidad asociada con la covariable tanto de la variabilidad no explicada como de la variabilidad del efecto del tratamiento.

En la Tabla 2.9 se muestra la estructura de la tabla de datos de partida para el ACOVAR unifactorial con una sola covariable.

Tabla 2.9. Matriz de datos para un ACOVAR unifactorial con una sola covariable.

	Grupo 1		Grupo 2		...	Grupo k	
	variable	covariable	variable	covariable		variable	covariable
Individuo 1	y_{11}	x_{11}	y_{12}	x_{12}	...	y_{1k}	x_{1k}
Individuo 2	y_{21}	x_{21}	y_{22}	x_{22}	...	y_{2k}	x_{2k}
·	·	·	·	·		·	·
·	·	·	·	·		·	·
·	·	·	·	·		·	·
Individuo n	y_{n1}	x_{n1}	y_{n2}	x_{n2}	...	y_{nk}	x_{nk}
Sumas	$\sum_{i=1}^n y_{i1}$	$\sum_{i=1}^n x_{i1}$	$\sum_{i=1}^n y_{i2}$	$\sum_{i=1}^n x_{i2}$...	$\sum_{i=1}^n y_{ik}$	$\sum_{i=1}^n x_{ik}$
Medias	$\bar{y}_{.1}$	$\bar{x}_{.1}$	$\bar{y}_{.2}$	$\bar{x}_{.2}$...	$\bar{y}_{.k}$	$\bar{x}_{.1}$

El modelo estadístico subyacente al ACOVAR es el siguiente:

$$y_{ij} = \mu + \alpha_j + \beta x_{ij} + \epsilon_{ij}$$

donde y_{ij} es la puntuación en la variable dependiente del individuo i del grupo j , μ es la media poblacional, α_j es el efecto del valor de la variable de agrupación que define el grupo j , β es el coeficiente de regresión de la variable dependiente sobre la covariable, x_{ij} es la puntuación del individuo i del grupo j en la covariable y ϵ_{ij} es el residual.

La inclusión de la covariable en el modelo estadístico afectará al análisis en dos sentidos (Maxwell y Delaney, 1990):

1. La suma de cuadrados intragrupo se reducirá en una magnitud que dependerá de la correlación entre la variable dependiente y la covariable. Así, se conseguirá una reducción en la variabilidad no explicada con el correspondiente incremento en la sensibilidad del análisis para detectar el efecto del tratamiento.
2. Se ajustará la estimación de la suma de cuadrados del efecto del tratamiento. El tamaño de este ajuste dependerá de la correlación que haya entre la variable dependiente y la covariable, y de las diferencias que presenten los grupos experimentales en la covariable.

Las condiciones que se han de cumplir para poder aplicar un ACOVAR son las siguientes:

1. Los residuales son independientes y se distribuyen según un modelo de probabilidad normal, con media igual a cero e igual variancia. El ACOVAR es bastante robusto ante las violaciones de la normalidad y la homogeneidad de variancia, sobre todo si hay un número igual de sujetos por condición de tratamiento (Glass, Peckham y Sanders, 1972). Sin embargo, igual que ocurre en el AVAR, este procedimiento no es robusto ante la violación de la independencia.
2. Linealidad entre la variable dependiente y la covariable. Esto quiere decir que la relación entre la variable dependiente y la covariable es lineal para todas las poblaciones de tratamiento. La inspección visual de los diagramas de

dispersión para cada grupo de tratamiento podrá darnos una primera información acerca de la plausibilidad de este supuesto.

3. Homogeneidad de la regresión. Esto quiere decir que el coeficiente de regresión es constante para todas las poblaciones de tratamiento. La condición se cumplirá si las pendientes de las rectas de regresión para todos los grupos de tratamiento son homogéneas en la población; es decir, si las rectas de regresión son paralelas. El cumplimiento de esta condición implica que no hay interacción entre la variable independiente y la covariable, por lo que el efecto de la variable manipulada no dependerá de los valores que cada sujeto obtenga en la covariable.
4. Fiabilidad en la medida de la covariable. La covariable se ha de medir sin error, puesto que, en caso contrario, se producen sesgos en la estimación de las pendientes de regresión de los grupos, aumentando la probabilidad de cometer error tipo I.

2.3.4. Diseños no experimentales

2.3.4.1. Definición y características

Comenzaremos este epígrafe ofreciendo algunas definiciones del concepto de diseño no experimental, y, a partir de ellas, expondremos las principales características que presenta este tipo de diseño.

Pedhazur y Pedhazur Schmelkin (1991) indican que los experimentos, los cuasi-experimentos y los no experimentos *se diferencian por la presencia o ausencia de (a) manipulación de la variable independiente, y (b) aleatorización. En un experimento tanto la manipulación como la aleatorización están presentes; en un cuasi-experimento, la manipulación está presente pero no la aleatorización; en un no experimento tanto la manipulación como la aleatorización están ausentes* (p. 304). Estos mismos autores puntualizan que los diseños no experimentales han recibido distintas denominaciones y que, por el momento, no hay consenso respecto al uso de una u otra. Tales denominaciones son: *investigación correlacional, investigación de encuestas o investigación observacional*.

Kerlinger y Lee (2002) nos dicen que *la investigación no experimental es la búsqueda empírica y sistemática en la que el científico no posee control directo de*

las variables independientes, debido a que sus manifestaciones ya han ocurrido o a que son inherentemente no manipulables. Se hacen inferencias sobre las relaciones entre las variables, sin intervención directa, de la variación concomitante de las variables independiente y dependiente (p. 504). Cabe subrayar en esta definición que la razón por la que no se manipula la variable independiente en la investigación no experimental es que resulta *imposible* hacerlo.

Arnau (1995a) utiliza el término investigación no experimental *para denominar genéricamente a un conjunto de métodos y técnicas de investigación distinto de la estrategia experimental y cuasi-experimental* (p. 35). Destaca que en este tipo de investigaciones no hay ni manipulación de la variable independiente ni aleatorización en la formación de los grupos.

A partir de las definiciones anteriores podemos elaborar un listado con las principales características de los diseños no experimentales. Son las siguientes:

1. No hay manipulación de la variable independiente, bien sea porque se trate de una variable que ya ha acontecido, bien sea porque se trate de una variable que por su propia naturaleza o por cuestiones éticas no pueda manipularse de forma activa. Se incluyen, pues, en este grupo de variables todas aquellas que recogen características propias de los individuos.
2. No hay asignación aleatoria de los sujetos a las condiciones de tratamiento. Es decir, los grupos no se forman aleatoriamente, por lo que no queda garantizada su equivalencia inicial.
3. Los datos simplemente se recolectan y luego se interpretan (Kerlinger y Lee, 2002), puesto que no se interviene de forma directa sobre el fenómeno.
4. Se estudian los fenómenos tal y como ocurren de forma natural. Esta característica hace que los diseños de tipo no experimental se utilicen principalmente en investigación aplicada.
5. De las características anteriores se desprende que el diseño no experimental no permitirá establecer relaciones causales inequívocas.

Si bien algunas de las características anteriores podrían desalentar al investigador de recurrir a un diseño no experimental vamos a presentar ahora algunas circunstancias en las cuales se deberá utilizar este tipo de diseño. En primer lugar, existen muchos problemas de investigación para los cuales no resulta apropiado un diseño experimental: un investigador puede estar interesado en

recoger información sobre las actitudes o creencias de los individuos con una finalidad puramente descriptiva o puede querer observar el fenómeno sin intervenir para no desvirtuarlo. En segundo lugar, nos encontramos con una buena parte de investigaciones que se realizan en el ámbito de la Psicología en las cuales el objetivo es estudiar características de las personas que no son susceptibles de manipulación experimental —por ejemplo, el cociente intelectual o el nivel de extraversión—. Esto hace que el investigador no pueda provocar el fenómeno que quiere estudiar, y, en consecuencia, que no pueda recurrir ni a un diseño experimental ni a un diseño cuasi-experimental. En tercer lugar, puede ocurrir que aunque la variable de interés se pueda manipular experimentalmente haya razones de tipo ético que lo impidan. Por ejemplo, no se podrá manipular una variable independiente si esta manipulación puede producir daños físicos o mentales a los participantes. Por último, hay ocasiones en las cuales no se podrá llevar a cabo un experimento verdadero por razones de tipo financiero, administrativos, o simplemente porque supondría excesivos inconvenientes para los participantes en la investigación.

2.3.4.2. Clasificación de los diseños no experimentales

No existe demasiado consenso acerca de cómo se han de clasificar los diseños no experimentales, por lo que vamos a presentar aquí algunas propuestas.

Kirk (1995) afirma que *la investigación científica se lleva a cabo con los siguientes objetivos: (1) explorar, (2) describir o clasificar, (3) establecer relaciones y (4) establecer causalidad. A lo largo de los años, los investigadores han desarrollado una variedad de estrategias de investigación para alcanzar estos objetivos: el experimento, el cuasi-experimento, la encuesta, el estudio de casos y la observación natural* (p.5). Más adelante añade otras estrategias de investigación a esta lista: los estudios *ex post facto*, los estudios retrospectivos y prospectivos, los estudios estáticos y longitudinales, y los estudios de series de tiempo y de sujeto único. Si bien este autor no menciona de forma explícita los diseños no experimentales, podemos, por exclusión, incluir en esta categoría todas aquellas estrategias de investigación distintas del experimento y del cuasi-experimento.

Smith y Davis (2003), siguiendo la propuesta de Kirk (1995), incluyen bajo la etiqueta de métodos no experimentales las siguientes estrategias de investigación: métodos descriptivos —se caracterizan porque se limitan a observar y documentar diversos aspectos de una situación o de un fenómeno que ocurre de forma natural

(dentro de estos métodos incluyen los estudios de casos y la observación natural)—; la investigación cualitativa —que según Creswell (1994) es *un proceso de indagación para comprender un problema humano o social basado en un complejo, figura holística, formado con palabras, recogiendo opiniones detalladas de los informantes y llevado a cabo en un ambiente natural* (p. 2)—; los estudios correlacionales —cuyo objetivo es establecer relaciones entre variables—; los estudios *ex post facto* —que se caracterizan porque las variables que se estudian se seleccionan después de que hayan ocurrido—; y las encuestas.

Polit y Hungler (2000) diferencian dentro de los diseños no experimentales entre la *investigación correlacional* y la *investigación descriptiva*. La primera se caracteriza porque su objetivo es estudiar las asociaciones entre las variables implicadas en la investigación. Es decir, se interesa en estudiar si las variaciones registradas en una variable se relacionan con las variaciones registradas en otra. La segunda se caracteriza porque se limita a observar un fenómeno que ocurre de forma natural. Dentro de la investigación correlacional se diferencia entre los *diseños retrospectivos* y los *diseños prospectivos*. Los primeros son investigaciones en las que un determinado fenómeno actual se vincula con otros cuya incidencia tuvo lugar antes del inicio del estudio. Esto quiere decir que el investigador centra su atención en un resultado presente e intenta arrojar luz sobre los antecedentes que lo han causado. Muchos estudios epidemiológicos e investigaciones médicas tienen carácter retrospectivo. Los diseños prospectivos, por su parte, se inician con el registro de los presuntos factores causales y prosiguen a lo largo del tiempo hasta producirse el efecto esperado. En general, los estudios prospectivos son más costosos que los estudios retrospectivos, y ésta es la razón por la que se utilizan con menos frecuencia. Dentro de la investigación descriptiva, Polit y Hungler distinguen entre la *observación natural* y las *encuestas*.

Arnau (1995a) clasifica los diseños no experimentales en función del conjunto de métodos de recogida de datos, y diferencia entre los *métodos observacionales* y los *métodos de encuesta*.

La metodología observacional *es una estrategia particular del método científico que se propone la cuantificación del comportamiento espontáneo que ocurre en situaciones no preparadas, implicando para su consecución el cumplimiento de una serie ordenada de etapas* (Anguera, 1990, p. 126). León y Montero (1997) definen la observación científica como un *plan de recogida de datos en el que el investigador registra, mediante un código más o menos estructurado, el desarrollo*

de los fenómenos de su interés. La situación en la que esto se hace puede ser natural o artificial (estructurada o construida por el investigador) (p. 67). Arnau (1995a) diferencia entre tres tipos de diseños de recogida de datos dentro de la metodología observacional: la *observación natural*, el *estudio de casos* y los *auto-informes*.

La observación natural implica observar individuos o eventos en su ambiente natural sin utilizar intervenciones manipulativas o técnicas de medida que podrían alterar ese ambiente (Kirk, 1995). Se trata de una forma de investigación pasiva en el sentido de que el individuo observado determina los eventos disponibles para ser registrados. El investigador es un registrador no intrusivo de los eventos que acontecen. En este caso, el investigador debe focalizar su atención en un número finito de eventos que deben estudiarse mediante la observación, ha de elaborar un sistema de categorías conductuales y ha de registrar las frecuencias y/o duración de las conductas. Señala Kirk que la observación natural es uno de los métodos de investigación más antiguos; pensemos, por ejemplo, en la investigación en astronomía, en las observaciones naturales de Charles Darwin o los estudios de chimpancés en Tanzania de Jane Goodall. Como estrategia de investigación la observación natural tiene dos claras ventajas respecto a situaciones más controladas: (1) ganan validez ecológica, puesto que el fenómeno se estudia en situación natural, y (2) se evitan los efectos reactivos ante la intervención del experimentador —es decir, las reacciones antinaturales—, puesto que los sujetos no se dan cuenta de que están siendo observados. Por desgracia la observación natural también presenta inconvenientes: (1) nos permite describir qué ocurre pero no responde a la cuestión de por qué ocurre, y (2) tampoco nos permite dar respuesta a la cuestión: ¿qué pasaría si...?

El estudio de casos se refiere a la observación exhaustiva de un solo individuo en situaciones no controladas (Kazdin, 2001). Este individuo suele ser un paciente, puesto que este tipo de estudios se han desarrollado principalmente en el ámbito clínico, del cual se recopila información de forma anecdótica y no sistemática. Esto hace que este tipo de estudios tengan muchos problemas, por lo que se refiere a su validez, si los comparamos con la investigación controlada. A pesar de los múltiples problemas que presenta este tipo de estudio, Kazdin (2001) señala sus virtudes: (1) ha servido de fuente de ideas e hipótesis acerca del comportamiento y desarrollo humanos; (2) con frecuencia ha sido el origen del desarrollo de técnicas terapéuticas; (3) permite el estudio de fenómenos extraños o poco frecuentes, que

difícilmente podrían estudiarse con un diseño de grupos; (4) es útil para proporcionar contraevidencias de las nociones que se consideran universalmente aplicables; y (5) tiene un valor persuasivo y motivacional, porque proporcionan demostraciones contundentes que hacen concreto y agudo lo que de otra manera podría servir como un principio abstracto.

El autoinforme consiste en *la simple observación de los pensamientos, impresiones o sentimientos propios del observador, mediante introspección* (Arnau, 1995a, p. 40). Así pues se produce una situación en la cual un mismo individuo actúa como observador y observado, con todos los problemas de validez que esto supone. Se trata de uno de los métodos más antiguos de la Psicología, aunque en la actualidad se ha abandonado, puesto que no se considera un medio propio de la investigación científica.

La metodología selectiva, o de encuesta, se define de dos formas distintas en la literatura sobre métodos de investigación (Martínez, 1995). Por una parte, la encuesta se define como *una investigación en la que se recoge información de forma sistemática, sin que el investigador lleve a cabo manipulaciones ni intervenciones* (Abramson, 1990, p. 10). Se entiende aquí la encuesta en un sentido amplio, dado que incluiría toda la investigación no experimental. Por otra parte, se define como *la recogida sistemática de datos de poblaciones o de muestras de poblaciones por medio de entrevistas personales u otros instrumentos de recogida, especialmente cuando se refieren a grupos de personas amplios o dispersos* (Dalenius, 1988, p. 17). Esta segunda definición de la metodología de encuestas es mucho más restrictiva que la anterior y es la que adoptan la mayoría de autores (por ejemplo, Arnau, 1995a). En esta misma línea, Gómez (1990) apunta que la metodología selectiva se caracteriza porque *intenta obtener información cuantitativa de la población —ya sea en términos descriptivos o de relación entre variables medidas— utilizando diseños que controlen de modo externo las condiciones de producción de la conducta mediante la adecuada selección de las unidades de análisis y la sistematización de la recogida de información* (p. 239).

Si entendemos la metodología de encuestas en un sentido restrictivo, éstas son sus principales características (Fowler, 1993): (1) Ausencia de manipulación o intervención por parte del investigador; (2) Su objetivo es obtener datos cuantitativos para describir algún aspecto de la población bajo estudio; (3) La información se obtiene preguntando a las personas, bien sea a través de cuestionarios o bien sea a través de entrevistas; (4) Se trabaja con una muestra

representativa de la población que se selecciona mediante algún diseño muestral de tipo probabilístico.

Martínez (1995) presenta dos clasificaciones de las encuestas. La primera diferencia entre descriptivas y analíticas. Las encuestas descriptivas son aquellas en las que se obtiene información de una sola muestra con objeto de estudiar un determinado fenómeno. Las encuestas analíticas son aquellas en las que se persigue estudiar las diferencias entre subgrupos de una población —en este caso interesa estudiar la relación o asociación entre variables—. La segunda clasificación diferencia entre encuestas transversales, longitudinales y de cohortes. La encuesta transversal recoge información sobre alguna característica de interés en la población en un único momento temporal. La encuesta longitudinal estudia la característica de interés en distintos momentos temporales con objeto de estudiar su evolución. La encuesta de cohortes estudia cohortes de sujetos longitudinalmente y además añade otras nuevas para poder llevar a cabo comparaciones.

2.3.4.3. Análisis de datos para este tipo de diseños

Dada la diversidad de métodos de los que se vale la investigación no experimental, resulta muy difícil presentar una síntesis de las técnicas de análisis de datos que se utilizan en ésta. Hay una buena parte de técnicas que son propias del diseño de encuestas, que constituyen, por sí mismas, materia de otras asignaturas de la licenciatura de Psicología, por lo que no consideramos adecuado incorporarlas a este texto.

REFERENCIAS BIBLIOGRÁFICAS

- Abramson, J. H. (1990). *Survey methods in community medicine*. London: Churchill Livingstone.
- American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Autor.
- Anderson, S., Auquier, A., Hauchk, W.W., Oakes, D, Vandeale, W., y Weisberg, H. (1980). *Statistical methods for comparative studies*. New York: John Wiley.
- Anguera, M. T. (1990). La metodología observacional. En J. Arnau, M. T. Anguera y J. Gómez (Eds.). *Metodología de la investigación en Ciencias del Comportamiento* (pp. 7-122). Murcia: Secretariado de publicaciones de la Universidad de Murcia.
- Arnau, J. (1984). *Diseños experimentales en psicología y educación. Vol. II*. Mexico, D.F.: Trillas.
- Arnau, J. (1989). Metodología de la investigación y diseños. En J. Mayor y J. L. Pinillos (Eds.), *Tratado de psicología general. Vol I: Teoría, historia y método* (Coords. J. Arnau y H. Carpintero) (pp. 581-616). Madrid: Alambra.
- Arnau, J. (1990a). Metodología experimental. En J. Arnau, M. T. Anguera, y J. Gómez, *Metodología de la investigación en ciencias del comportamiento*. Murcia: Secretariado de Publicaciones de la Universidad de Murcia.
- Arnau, J. (1990b). *Diseños experimentales multivariables*. Madrid: Alianza.
- Arnau, J. (1994). Diseños experimentales de caso único. En R. Fernández Ballesteros (Ed.), *Evaluación conductual hoy: Un enfoque para el cambio en psicología clínica de salud* (pp.40-86). Madrid: Pirámide.

- Arnau, J. (1995a). Metodología de la investigación psicológica. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología* (pp. 23-43). Madrid: Síntesis.
- Arnau, J. (1995b). Fundamentos metodológicos de los diseños experimentales de sujeto único. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología* (pp. 179-193). Madrid: Síntesis.
- Arnau, J. (1995c). Diseños experimentales de sujeto único. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología* (pp. 179-193). Madrid: Síntesis.
- Arnau, J. (1995d). Análisis estadístico de datos para los diseños de sujeto único. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología* (pp. 179-193). Madrid: Síntesis.
- Arnau, J. (1995e). Metodologies quantitatives en la investigació psicològica. En J. Arnau (Ed.), *Mètodes, dissenys i tècniques en investigació psicològica*. Barcelona: Universitat Oberta de Catalunya.
- Arnau, J. (1997). *Diseños de investigación aplicados en esquemas*. Barcelona: Edicions U.B.
- Arnau, J. (Ed.) (2001). *Diseños de series temporales: técnicas de análisis*. Barcelona: Edicions de la Universitat de Barcelona.
- Arnau, J. y Bono, R. (2003). *Diseños de investigación aplicados (CD-ROM)*. Barcelona: Edicions de la Universitat de Barcelona.
- Arnau, J. y Bono, R. (2006). *Evaluación de las competencias de la materia Métodos, Diseños y Técnicas de Investigación en Psicología*. Comunicación presentada en el 4º Congreso Internacional "Docencia Universitaria e Innovación", Barcelona, España.
- Ato, M y Vallejo, G. (2007). *Diseños experimentales en Psicología*. Madrid: Pirámide.
- Ato, M. (1991). *Investigación en ciencias del comportamiento I: Fundamentos*. Barcelona: PPU.
- Ato, M. (1995a). Concepto básicos. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología* (pp. 45-70). Madrid: Síntesis.
- Ato, M. (1995b). Tipología de los diseños cuasi-experimentales. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología* (pp. 245-269). Madrid: Síntesis.

- Ato, M. (1995c). Análisis estadístico II: Diseños con variable de asignación conocida. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología* (pp. 305-320). Madrid: Síntesis.
- Ato, M. (1995d). Análisis estadístico I: Diseños con variable de asignación no conocida. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología* (pp. 271-304). Madrid: Síntesis.
- Baer, D. M., Wolf, M. M. y Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1, 91-97
- Balluerka, N. (2001). Diseños de sujeto único. En J. Arnau (Ed.) *Diseños de series temporales: técnicas de análisis* (pp. 235- 290). Barcelona: Edicions de la Universitat de Barcelona.
- Barlow, D. H. y Hersen, M. H. (1984). *Single-case experimental designs: strategies for studying behavior change* (2nd ed.). Oxford, UK: Pergamon Press.
- Bickman, L. y Rog, D. J. (1998). *Handbook of applied social research methods*. Thousand Oaks: Sage.
- Bono, R. (1997). *Proyecto docente de la asignatura Diseños de Investigación Aplicados*. Barcelona: Autor.
- Bordens, K. S. y Abbot, B. B. (1988). *Research design and methods: A process approach*. New York, NY: Mayfield Publishing Company.
- Box, G. E. P. y Jenkins, G. M. (1970). *Time series análisis: Forecasting and control*. San Francisco, CA: Holden-Day.
- Box, G. E. P. y Tiao, G. C. (1965). A change in level of a nonstationary time series. *Biometrika*, 52, 181-192.
- Box, G. E. P., Hunter, W. G., y Hunter, J. S. (1978). *Statistics for experimenters. An introduction to design, data analysis, and model building*. New York, NY: John Wiley and Sons.
- Campbell, D. T. y Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin Company.
- Chassan, J. B. (1967). *Research design in clinical psychology and psychiatry*. New York: Appleton-Century-Crofts.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

- 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.
- Cozby, P. C. (1993). *Methods in behavioral research* (5th ed.). Mountain View, CA: Mayfield Publishing Company.
- Creswell, J. W. (1994). *Research design: Qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 30, 116-127.
- Dalenius, T. (1988). A first course in survey sampling. En P. R. Krishnaiah y C. R. Rao (Eds.). *Handbook of statistics, Vol 6. Sampling*. Amsterdam: North Holland.
- Diener, D. y Crandall, R. (1979). *Ethics in social and behavioral Research*. Chicago: University of Chicago Press.
- Dwyer, J. H. (1983). *Statistical models for the social and behavioral sciences*. New York: Oxford University Press.
- Edgington, E. S. (1967). Statistical inference from N=1 experiments. *Journal of Psychology*, 65, 195-199.
- Edgington, E. S. (1972). N=1 experiments: Hypothesis testing. *Canadian Psychologists*, 13, 121-135.
- Edgington, E. S. (1995). *Randomization tests* (3^a ed.). New York: Marcel Dekker.
- Edwards, A. L. (1985). *Experimental design in psychological research* (5th ed.). New York: Harper and Row.
- Field, A. y Hole, G. J. (2003). *How to design and report experiments*. London: Sage.
- Fisher, R. A. (1932). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.
- Fisher, W. W., Kelley, M. E. y Lomas, J. E. (2003). Visual aids and structured criteria for improving visual inspection and interpretation of single-case design. *Journal of Applied Behavior Analysis*, 36, 387-406.
- Fowler, F. F. (1993). *Survey research methods*. Newbury Park, CA: Sage.
- Franklin, R. D., Allison, D. B. y Gorman, B. S. (1996). *Design and analysis of single-case research*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Franklin, R. D., Gorman, B. S., Beasley, T. M. y Allison, D. B. (1996). Graphical display and visual analysis. En R. D. Franklin, D. B. Allison y B. S. Gorman,

- Design and analysis of single-case research (pp. 119-158). Mahwah, NJ: Lawrence Erlbaum Associates.
- Geisser, S. y Greenhouse, S. W. (1958). An extension of Box's results on the use of the F distribution in multivariate analysis. *Annals of Mathematical Statistics*, 29, 885-891.
- Gentile, J. R., Roden, A. H. y Klein, R. D. (1972). An analysis of variance model for the intrasubject replication design. *Journal of Applied Behaviour Analysis*, 5, 193-198.
- Glass, G. V., Peckham, P. D. y Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the analysis of variance and covariance. *Review of educational research*, 42, 237-288.
- Glenn, N. D. (1977). *Cohort analysis*. Newbury Park, CA: Sage.
- Gómez, J. (1990). Metodología de encuesta por muestreo. En J. Arnau, M. T. Anguera y J. Gómez (Eds.), *Metodología de la investigación en Ciencias del Comportamiento*. Murcia: Secretariado de publicaciones de la Universidad de Murcia.
- Gorman y Allison (1996). Statistical alternatives for single-case designs. En R. D. Franklin, D. B. Allison y B. S. Gorman (Eds.), *Design and analysis of single-case design* (pp.157-214). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hayes, S. C. (1981). Single case experimental design and empirical clinical practice. *Journal of Consulting and Clinical Psychology*, 49, 193-211.
- Hays, W. L. (1994). *Statistics* (5th ed.). Belmont, CA: Thomson Learning.
- Hedrick, T. E., Bickman, L. y Rog, D. J. (1993). *Applied research design. A practical guide*. Newbury Park, CA: Sage.
- Hernández, F. (1997). La investigación experimental. En L. Buendía, M. P. Colás y F. Hernández (Eds.), *Métodos de investigación en psicopedagogía*. México: McGraw-Hill.
- Hersen, M. y Barlow, D. H. (1976). *Single case experimental designs: strategies for studying behavior change*. New York: Pergamon Press.
- Huynh, H. y Feldt, L. S. (1976). Conditions under which mean square ratios in repeated measurement designs have exact F -distributions. *Journal of the American Statistical Association*, 65, 1582-1589.
- Judd, C. M. y Kenny, D. A. (1981). *Estimating the effects of social interventions*. Cambridge: Cambridge University Press.

- Kazdin, A. E. (1982). *Single-case research designs: methods for clinical and applied settings*. New York: Oxford University Press.
- Kazdin, A. E. (2001). *Métodos de investigación en psicología clínica* (3ª ed.). Mexico: Prentice Hall.
- Kazdin, A. E. y Hartmann, D. P. (1978). The simultaneous-treatment design. *Behavior Therapy*, 9, 912-922.
- Kennedy, C. H. (2005). *Single-case designs for educational research*. Boston, MA: Pearson Education, Inc.
- Kenny, D. A. (1975). 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: Wiley and sons.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs: Prentice-Hall Inc.
- Keppel, G. y Zedeck, S. (1989). *Data analysis for research. Analysis of variance and multiple regression/correlation approaches*. San Francisco, CA: Freeman and Company.
- Kerlinger, F. M. (1986). *Foundations of behavioural research* (3rd ed.). New cork: Holt, Rinehart and Winston.
- Kerlinger, F. N. y Lee, H. B. (2002). *Investigación del comportamiento. Métodos de investigación en ciencias sociales* (4ª ed.). Mexico: McGraw-Hill.
- Keselman, H. J., Keselman, J. C. y Lix, L. M. (1995). The analysis of repeated measurements: Univariate tests, multivariate tests, or both? *British Journal of Mathematical and Statistical Psychology*, 48, 319-338.
- Kirk, R. E. (1995). *Experimental design: procedures for the behavioral sciences* (3rd ed.). Belmont, CA: Brooks/Ciole Publishing.
- Knapp, T. J. (1983). 'Behavior' analysts visual appraisal of behavior change in graphic display. *Behavioral Assessment*, 5, 155-164.
- Kratochwill, T. R. (1992). Single-case research design and analysis: An overview. En T. R. Kratochwill y J. R. Levin (Eds.), *Single-case research design and analysis: New directions for Psychology and Education* (pp. 1-14). Hillsdale, NJ: Erlbaum Lawrence Associates.
- Kratochwill, T. R. y Brody, G. H. (1978). Single subject designs: A perspective on the controversy employing statistical inference and implications for research and training in behaviour modification. *Behavior Modification*, 2, 291-307.

- Kratochwill, T. R., Mott, S. E. y Dodson, C. L. (1984). Case study and single-case research in clinical and applied psychology. En A. S. Bellack y M. Hersen (Eds.), *Research methods in clinical psychology*. New York: NY: Pergamon Press.
- Leitenberg, H. (1973). The use of single-case methodology in psychotherapy research. *Journal of Abnormal Psychology*, 82, 87-101.
- León, O. y Montero, I. (1997). *Diseño de investigaciones. Introducción a la lógica de la investigación en Psicología y Educación* (2ª ed.). Madrid: McGraw-Hill.
- Marasinghe, M. G., Meeker, W. Q., Cook, D. y Shin, T. (1996). Using graphics and simulation to teach statistical concepts. *The American Statistician*, 50, 342-351.
- Martínez, R. (1995). El método de encuestas por muestreo: conceptos básicos. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología* (pp. 385-431). Madrid: Síntesis.
- Mauchly, J. W. (1940). Significance test for sphericity of a normal n -variate distribution. *Annals of Mathematical Statistics*, 29, 204-209.
- Maxwell, S. E. y Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective*. Belmont: Wadsworth Publishing Company.
- Maxwell, S. E., O'Callaghan, M. F. y Delaney, H. D. (1993). Analysis of covariance. En L. K. Edwards (Ed.), *Applied analysis of variance in behavioral science* (pp. 63-104). New York: Marcel Dekker, Inc.
- McNamara, J. R. y MacDonough, T. S. (1972). Some methodological considerations in the design and implementation of behavior therapy research. *Behavior Therapy*, 3, 361-378.
- Mead, R. (1988). *The design of experiments. Statistical principles for practical application*. Cambridge: Cambridge University Press.
- Meltzoff, J. (2000). *Crítica a la investigación: Psicología y campos afines*. Madrid: Alianza Editorial.
- Michell, J. (1985). Measurement scales and statistics. A clash of paradigms. *Psychological bulletin*, 100(3), 398-407.
- Miller, D. C. (1991). *Handbook on research design and social measurement* (5th ed.). Newbury Park, CA: Sage.
- Millon, T. y Diefenhaus, H. I. (1972). *Research methods in psychopathology*. New York: John Wiley.

- Moreno, R y López, J. (1985). *Análisis metodológico de investigaciones experimentales en psicología*. Barcelona: Alamex.
- Morley, S. y Adams, M. (1991). Graphical analysis of single-case time series data. *British Journal of Clinical Psychology*, 30, 97-115.
- Parsonson, B. S. y Baer, D. M. (1978). The analysis and presentation of graphic data. En T. R. Kratochwill (Ed.), *Single-subject research: Strategies for evaluating change* (pp. 101-165). New York: Academic Press.
- Pedhazur, E. J. y Pedhazur Schmelkin, L. P. (1991). *Measurement, design, and analysis. An integrated approach* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Pereda, S. (1987). *Psicología experimental: I. Metodología*. Madrid: Pirámide.
- Pizarro, F. (1986). *Aprender a razonar*. Madrid: Editorial Alhambra.
- Poling, A. Methot, L. L. y LeSage, M. G. (1995). *Fundamentals of behaviour analytic research*. New York, NY: Plenum Press.
- Polit, D. F. y Hungler, B. P. (2000). *Investigación científica en ciencias de la salud* (6^a ed.) Mexico: McGraw-Hill.
- Ramsey, P. H. (1993). Multiple comparisons of independent means. En K. L. Edwards (Ed.), *Applied analysis of variance in behavioural science* (pp. 25-62). New York: Marcel Dekker, Inc.
- Roscoe, J. T. (1975). *Fundamental research statistics for the behavioural sciences* (2nd ed.). New York: Holt, Rinehart and Winstons.
- Rossi, P. H. y Wright, J. D. (1985). Evaluation research: an assessment. En L. H. Aiken y B. H. Kehrner (Eds.), *Evaluation studies review annual*, vol 10. Beverly Hills, CA: Sage.
- Rubin, A., Rosebery, A., Bruce, B. y DuMouchel, W. (1988). Getting an early start: Using interactive graphics to teach statistical concepts in high school. En *Proceedings of the Statistical Education Section, American Statistical Association*. New Orleans: American Statistical Association.
- Ryder, N. B. (1965). The cohort as a concept in the study of social change. *American review of sociology*, 30, 843-861.
- Saville, B. (2007). *A guide to teaching research methods in Psychology*. Blackwell Publishing.
- Seco, M. (1999). *Diccionario del español actual*. Madrid: Aguilar.

- Shadish, W. R., Cook, T. D. y Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company.
- Shapiro, M. B. (1961). The single case in fundamental clinical psychological research. *British Journal of Medical Psychology*, 34, 255-263.
- Shapiro, M. B. (1966). The single case in clinical-psychological research. *Journal of General Psychology*, 74, 3-23.
- Shine, L. C. y Bower, S. M. (1971). A one-way analysis of variance for single-subject designs. *Educational and Psychological Measurement*, 31, 105-113.
- Sidman, M. (1952). A note on functional relations obtained from group data. *Psychological bulletin*, 49, 263-269.
- Sidman, M. (1960). *Tactics of scientific research: evaluating experimental data in psychology*. Nueva York, NY: Basic Books.
- Smith, R. A. y Davis, S. F. (2003). *The psychologist as detective: An introduction to conducting research in psychology* (3rd ed.). Upper Saddle River, NJ: Pearson.
- Solanas, A., Salafranca, L., Fauquet, J. y Núñez, M. I. (2005). *Estadística descriptiva en ciencias del comportamiento*. Madrid: Thomson.
- Spector, P. E. (1981). *Research design*. London, UK: Sage Publications Quantitative Applications in the Social Sciences series.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677-680.
- Stevens, S. S. (1951) *Handbook of experimental psychology*. New York: Wiley.
- Trochim, W. M. K. (1984). *Research design for program evaluation: The regression-discontinuity approach*. Beverly Hills, CA: Sage.
- Trochim, W. M. K. (2006). *The research methods knowledge base* (2nd ed.). Disponible en <http://www.atomicdog.com>.
- Tukey, J. (1977). *Exploratory data analysis*. Cambridge, MA: Addison-Wesley.
- Ulman, J. D. y Sulzer-Azaroff, B. (1975). Multielement baseline design in educational research. En E. Ramp y G. Semb (Eds.), *Behavior analysis: Areas of research and application* (pp. 377-391). Englewood Cliffs, NJ: Prentice-Hall.
- Vallejo, G. (1995a). Análisis de los diseños de series de tiempo interrumpidas. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología* (pp. 335-351). Madrid: Síntesis.
- Vallejo, G. (1995b). Problemas inferenciales asociados con el uso de diseños de series de tiempo interrumpidas. En M. T. Anguera, J. Arnau, M. Ato, R.

- Martínez, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología* (pp. 353-382). Madrid: Síntesis.
- 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 educativa*. Buenos Aires: Paidós.
- Vasey, M. W. y Thayer, J. F. (1987). The continuing problem of false positives in repeated measures ANOVA in psychophysiology: a multivariate solution. *Psychophysiology*, 24(4). 479-486.
- Viader, M. (1996). *Diseños y análisis de experimentos en ciencias del comportamiento*. Barcelona: PPU.
- Visser, R. A. (1985). *Analysis of longitudinal data in behavioural and social research*. Leiden: DSWO Press.
- Wampold, B. E. y Furlong, M. J. (1981). The heuristics of visual inference. *Behavioral Assessments*, 3, 79-92.
- Wilkinson, L., y APA Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594.
- Winer, B. J., Brown, D. R. y Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.