

Diseños de $N = 1$

LOS denominados diseños de $N = 1$ o de *caso único* constituyen una categoría especial que históricamente surge en un marco concreto de investigación y que, aún en la actualidad, están restringidos principalmente a un ámbito reducido de aplicación. No por ello hay que asignarles menos mérito, pues al fin y al cabo cada método vale para lo que sus usuarios lo hacen valer. Digamos que el diseño de $N = 1$ tiene un uso muy localizado. Por ejemplo, en la revista *Journal of Applied Behavior Analysis*, el 89% de los estudios empíricos publicados en el 1989 eran experimentos de caso único.

El estatus científico del diseño ha sido cuestionado por algunos autores y otros no se ponen de acuerdo en donde situarlo. Unos lo consideran como *diseño aplicado*, otros como *experimental*, los terceros como *cuasi-experimental*... Por supuesto que existen ejemplos de aplicación de diseño de $N = 1$ que son perfectamente *experimentales* del mismo modo que también hay trabajos cuya metodología es *cuasi-experimental* (Edgington, 1987; Cook y Campbell, 1979).

Hay una serie de creencias incorrectas respecto a los diseños de $N = 1$ que crean confusión, destacando especialmente dos. En primer lugar, no hay que confundir los diseños de $N = 1$ con los *estudios de caso* tradicionales ya que en este último tipo de estudios el análisis se lleva a cabo a través de un único sujeto pero sus objetivos son exploratorios, no existiendo suficiente control de la situación (Kazdin, 1981). En segundo lugar, en ocasiones se asocian las investigaciones que utilizan un amplio número de sujetos con análisis estadísticos mientras que las que incluyen un único sujeto tienen que estar asociadas necesariamente a análisis visual de los resultados. Es un grave error. Los estudios de $N = 1$ pueden ser analizados con modelos estadísticos y los que incluyen un $N > 1$ pueden ser abordados con modelos no estadísticos como las técnicas de inspección visual (Kratowill, 1978; Martínez, 1986). El número de sujetos no determina un modelo de análisis.

Sin embargo, los diseños de $N = 1$ presentan problemas de validez *interna* y *externa*. Los problemas de validez *externa* pueden ser obvios desde el momento que se utiliza un sólo sujeto y la generalidad de los datos puede ser ampliamente cuestionada. En los diseños de $N = 1$ la “replicación” es la clave de la generalización. La replicación implica generar nuevos estudios que especifican claramente las condiciones de tratamiento y medida. La validez *externa* de los diseños de $N = 1$ depende así de la replicación

sistemática de los efectos en un amplio número de sujetos (Martínez, 1986). Barlow y Hersen (1984) describieron tres tipos distintos de replicación:

- *directa*: los mismos tratamientos con nuevos sujetos
- *sistemática*: cambio de las variables de interés como contexto, tipo de desajuste o trastorno...
- *clínica*: comprobación de paquetes de tratamiento con sujetos que presentan problemas conductuales semejantes

Es evidente, que el primer tipo de replicación tiene que ver con la fiabilidad de la investigación y, por tanto, se relaciona principalmente con la validez *interna* de la misma. En cambio, los otros dos tipos de replicación tienen que ver con la validez *externa*. Los autores últimamente citados llegan a afirmar que “...una serie de diseños de caso único con sujetos semejantes a los que se les aplica el mismo tratamiento original -replicación-, puede sobrepasar en tres o cuatro veces el diseño experimental con grupos de tratamiento / grupo de control” (pág. 57).

Otras amenazas contra la validez *interna*, por ejemplo, la *interacción selección por tratamiento*, son difíciles de controlar en este tipo de diseños: un sujeto particular con unas características individuales, puede reaccionar de forma específica al tratamiento, haciendo igualmente difícil la atribución causal que se pretende. En estos diseños, al igual que en los diseños *cuasi-experimentales*, hay que extremar las condiciones de control y tomar las precauciones necesarias para que factores tales como la historia, la maduración ... no afecten seriamente la validez.

Otro tipo de amenaza sería contra la validez *interna*, y que puede afectar también a la validez de *conclusión estadística*, procede del hecho mismo de la repetición de la secuencia *tratamiento – no tratamiento*. Cuando a un sujeto se le dan múltiples tratamientos de manera serial o el mismo tratamiento en múltiples etapas, ¿hay interferencia? ¿hay facilitación? ¿los efectos de un tratamiento dependen del orden en que se hayan administrado? ¿es diferente el efecto según sea en presencia o ausencia de un segundo elemento? De nuevo la salida airosa que, como solución general, se puede aportar parece venir de la replicación de secuencias alternativas.

Distintos diseños de N = 1

Estos diseños han recibido multitud de nombres: diseños de *caso único*, diseños de *series temporales*, diseños *intensivos*... y nosotros nos hemos quedado con el menos comprometedor de diseños de $N = 1$. Existen dos tipos de diseños especialmente útiles dentro de este ámbito: *diseño de inversión* y el *diseño de línea base múltiple*.

Un prototipo del *diseño de inversión* se puede representar simbólicamente de la siguiente manera:

$$A - B - A$$

donde *A* indica la primera fase de recogida de los datos, anterior a la aplicación del tratamiento y que describe la *línea de base* de conducta del sujeto, precisamente aquella sobre la que se quiere actuar. La letra *B* indica la fase de intervención, durante la cual tendrá lugar una nueva recogida de datos. Si los datos recogidos en esta segunda fase van en la dirección esperada, de nuevo introduciremos una fase de no-intervención (*A*), para comprobar si los datos se ajustan de nuevo al punto de partida (Kazdin, 1982). El término ‘inversión’ se refiere a la eliminación o retirada de la fase de tratamiento (*B*) realizándose posteriormente una nueva evaluación (*A*) de tal manera que si la fase *B* tuvo un efecto sobre la conducta objeto de estudio entonces un deterioro de los datos seguirá a su retirada.

Modelos más complejos añaden nuevas fases de intervención y no intervención. Así, por ejemplo:

$$A - B - A - B \quad \text{ó} \quad B - A - B - A$$

Si los cambios ocurren siempre que se introduce el tratamiento (*B*) y se genera un proceso de desaceleración o deterioro del cambio positivo al reintroducir la fase de no intervención (*A*), es lógico pensar que debe existir una cierta relación causal entre tratamiento y mejora.

La mayoría de estos diseños de *caso único* han sido aplicados por conductistas interesados por la solución de problemas de conducta, pero su uso no queda limitado a esta área de intervención. Las posibilidades no conductuales de estos diseños fueron demostradas hace tiempo por Truax y Carkhuff (1965) en una investigación basada en la terapia centrada en el cliente, aplicando un diseño del tipo *B-A-B*.

Los diseños de *caso único*, en principio ni están asociados ni están a favor de una orientación conductista. McCulloch, citado en Edgington (1987), por ejemplo los utilizó para investigar conductas cognitivas o fenómenos no-operantes que en nada cumplían los supuestos conductuales que habitualmente se exigen en la aplicación de estos diseños (que la variable dependiente sea estable antes de ser introducido el tratamiento, *-línea base estable-*, que sea fehacientemente demostrable la influencia de la variable independiente sobre la variable dependiente, y que los cambios de tratamiento sean rápidamente reflejados en la variable dependiente).

En ocasiones puede ocurrir que la variable tratamiento no puede ser retirada debido a limitaciones prácticas como presencia de efectos acumulativos que persisten después de la retirada del tratamiento o

consideraciones éticas que impiden su retirada (Barlow y Hersen, 1973). En estos casos, los *diseños de línea de base múltiple* aportan una solución analítica a este tipo de estudios. En estos diseños se identifican y miden un cierto número de respuestas o conductas con objeto de proporcionar *líneas de base* con las que se puedan evaluar los cambios alcanzados con el tratamiento. Se trata así de obtener diversas *líneas de base* y después comprobar de manera secuencial el efecto de la variable tratamiento en cada una de ellas. Se construyen por lo tanto diversos diseños *A – B* independientes que permiten dar fiabilidad al efecto del tratamiento ya que si dicho efecto se replica en diferentes puntos se puede garantizar en mayor medida que la intervención es la causante del cambio.

Existen tres tipos básicos de diseños de *línea de base múltiple*:

- *diseño de línea de base múltiple ‘entre conductas’*: un mismo sujeto es evaluado a lo largo de una variedad de conductas objetivo distintas e independientes.
- *diseño de línea de base múltiple ‘entre sujetos’*: en diferentes sujetos que están apareados y que se suponen están expuestos a las mismas condiciones ambientales se evalúa una misma conducta o tratamiento
- *diseño de línea de base múltiple ‘entre situaciones’*: un tratamiento específico se evalúa en un único sujeto a lo largo de situaciones distintas e independientes

Metodología de análisis

Un problema añadido que han planteado este tipo de diseños es el de su ‘análisis estadístico’. Son varias las propuestas que se han desarrollado (para una revisión en castellano véase Arnau, 1984; Kazdin, 1984; Martínez, 1986), destacando las *pruebas no paramétricas de aleatorización*, ampliamente defendidas por Edgington (1987), como especialmente útiles cuando el diseño de $N = 1$ cumple con los requisitos de ser un diseño experimental (conviene recordar que experimentación es igual a ‘manipulación’ y ‘aleatorización’). Además, los datos pueden ser abordados a través de la *prueba F* o diversas variaciones de ella, análisis de *series temporales*, la técnica de *división en mitades* o la *prueba de rangos R_n* , aplicándose esta última técnica cuando los datos se han obtenido en *diseños de línea base múltiple*. Nos vamos a detener especialmente en las *pruebas de aleatorización*.

La utilización de las pruebas paramétricas convencionales como *t* o *razón F* a través de diseños de *medidas repetidas* es considerada en general como inadecuada e inútil dada la dependencia serial que suelen presentar los datos de una investigación basada en un diseño donde las mediciones son realizadas a un mismo y único sujeto (Kratowill, Aldin, Demuth, Dawson, Panicucci, Arntson, McMurray, Hempstead, y Levin 1974; Huitema, 1985). Se considera que existe ‘dependencia serial’ en los datos, violándose así el supuesto de independencia de las observaciones, cuando

se produce *autocorrelación* entre los mismos de forma estadísticamente significativa.

Los valores de las pruebas estadísticas aplicadas se ven sesgados, de tal manera que el error de *Tipo I* es mayor cuando existen *autocorrelaciones positivas* entre los datos y menor cuando son *negativas* (Padia, 1973; tomado de Martínez, 1986) (véase *Tabla 1*).

La autocorrelación positiva disminuye espuriamente la varianza de error y por lo tanto amplía el nivel de error de *Tipo I*, haciendo la prueba estadística más liberal. Cuando la autocorrelación es negativa, la varianza de error aumenta también espuriamente, disminuyendo el error de *Tipo I*, siendo la prueba más conservadora (Sharpley, 1988).

Scheffé (1959) señaló que cuando la autocorrelación es de 0.30 el riesgo de error de *Tipo I* se incrementa de 0.05 a 0.12; incluso con autocorrelaciones menores como 0.20, el efecto de la autocorrelación se deja notar, incrementándose el error de *Tipo I* de 0.05 a 0.10.

En definitiva, a medida que la autocorrelación aumenta, el riesgo de error de *Tipo I* también aumenta rápidamente.

Solamente serán adecuadas las pruebas paramétricas cuando se compute *a priori* la autocorrelación de los datos y se compruebe que efectivamente no es significativa (Hartmann, 1974).

Si la autocorrelación es estadísticamente significativa es necesario aplicar otras pruebas estadísticas alternativas que obvien dicho fenómeno o no se vean afectadas por él.

Tabla 1 *Autocorrelación y Error de Tipo I*

<i>Autocorrelación</i>	$\alpha = 0.01$	$\alpha = 0.05$
0.50	1.3620	0.2585
0.40	0.0910	0.2006
0.20	0.0348	0.1074
0.00	0.0100	0.0500
-0.20	0.0016	0.0168
-0.40	0.0001	0.0026
-0.50	0.0000	0.0006

El estudio de la *autocorrelación* se lleva a cabo a través del cálculo de correlaciones entre los datos. Consideremos un ejemplo de cálculo de la autocorrelación.

Supuesto 1

Un investigador dispone de los siguientes datos correspondientes a dos fases o condiciones de su investigación (*condición A* y *condición B*). Decide comprobar si se ha producido una ‘dependencia serial’ entre los datos y para ello recurre al cálculo de la *autocorrelación* de los mismos comprobando su significación estadística. El *cálculo tiene que ser realizado por*

separado para cada condición y no de forma conjunta, computándose tantas autocorrelaciones como fases o condiciones tenga el estudio. Por lo tanto, en este caso es necesario calcular dos correlaciones de los datos: una para los datos de la *condición A* y otra para los datos de la *condición B*. Los datos obtenidos para la condición A y la condición B son los especificados en la siguiente tabla.

Recordar que el cómputo de la ‘autocorrelación’ tiene que ser realizado por separado para cada condición y no de forma conjunta. Trabajando con un nivel de probabilidad dado se considerará que no existe dependencia serial entre los datos si mantenemos el *modelo de la hipótesis nula o modelo restringido*.

Tabla 2 Datos del estudio

Condición A	Condición B
22	34
8	30
11	32
16	30
10	16
12	16
22	15
19	27
$\bar{X}_A = 15$	$\bar{X}_B = 25$

La ‘autocorrelación’ para cada condición ha sido calculada a través de un *coeficiente de correlación de Pearson* entre las columnas resultantes de ir combinando cada dato con el posterior (véase **Tabla 3**). La ‘autocorrelación’ por lo tanto permite conocer en qué grado las puntuaciones obtenidas en un momento predicen las obtenidas en un momento posterior.

Tabla 3 Datos del estudio para el cálculo de la autocorrelación

Condición A		Condición B	
22	8	34	30
8	11	30	32
11	16	32	30
16	10	30	16
10	12	16	16
12	22	16	15
22	19	15	27
Autocorrelación = -0.068 $p > 0.05$		Autocorrelación = 0.553 $p > 0.05$	

Se concluye que ni en la *condición A* ni en la *condición B* los datos presentan ‘dependencia serial’ estadísticamente significativa, siendo independientes los datos entre sí y no generando resultados espurios.

El cierto que la *inspección* o *análisis visual* de los datos puede dar información relevante en las fases exploratorias de una investigación facilitando la creación de hipótesis pero puede engañar al investigador si los datos presentan tendencias y dependencia serial dando lugar a resultados falsos. **El uso de procedimientos estadísticos que permitan verificar la presencia o no de autocorrelación es indispensable cuando se trabaja con datos generados secuencialmente por un mismo sujeto.**

Pruebas de aleatorización

En las *pruebas de aleatorización* la introducción del tratamiento y la retirada del mismo en vez de depender de la respuesta del sujeto, se hace de modo aleatorio. Los tratamientos (incluida la *línea base*) son aleatoriamente asignados a los diferentes momentos de aplicación, no dependiendo de las conductas del sujeto. De este modo el investigador evita que se generen ciertas tendencias de respuesta que probablemente hubieran surgido igualmente sin aplicar el tratamiento.

La manipulación aleatoria sirve para controlar el efecto de esas variables ocultas (*efecto de confundido*) y al mismo tiempo sienta las bases que hacen posible la aplicación de una prueba estadística posterior al análisis de los datos.