

Diseño experimental sin estadística

**Usos y restricciones en su
aplicación a las
ciencias de la conducta**

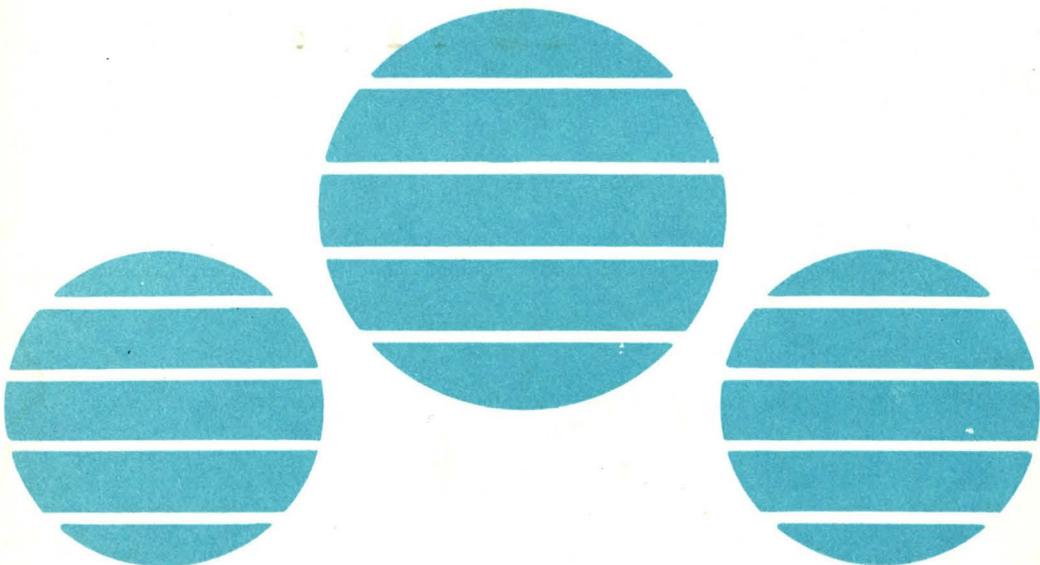
Luis Castro



**EDITORIAL
TRILLAS**

México, Argentina, España
Colombia, Puerto Rico, Venezuela

Revisión técnica: **GUSTAVO FERNANDEZ P.**
Profesor de tiempo completo
en la Facultad de Psicología de la
Universidad Nacional Autónoma de México.



**Diseño
experimental
sin
estadística**

Prólogo

Este libro no sostiene un enfoque antiestadístico; su título, sin embargo, puede inducir al error de considerarlo así. Varias personas han llamado mi atención a este respecto y me han hecho notar que, en casos extremos ya he sido catalogado como enemigo de la estadística. Nada más lejos de la verdad.

He considerado conveniente, por lo tanto, enfatizar aquí la tesis del libro, a saber: *es posible desarrollar y usar diversos diseños experimentales, independientemente de la estadística*. Esta última puede usarse en la etapa del análisis de los datos pero no es absolutamente indispensable entonces, y a menudo representa un obstáculo durante la etapa de planeación de la investigación. La estadística no tiene por qué determinar el arreglo de las condiciones experimentales que permitan contestar una pregunta de investigación o poner a prueba una hipótesis científica. Esta es una cuestión que se va a determinar por las consideraciones acerca del costo experimental, la capacidad informativa y otras características de los diseños alternativos que son de carácter lógico y metodológico, pero no estadístico.

Esta segunda edición incluye una nueva sección acerca de la estabilidad de la línea base en la investigación con sujetos humanos, así como algunas correcciones de organización y tipográficas; a más de una lista de referencias bibliográficas más amplia y actualizada que la de la primera edición.

En espera de que la obra cumpla adecuadamente con la función de proporcionar una guía para investigadores y estudiantes de la conducta en ámbitos experimentales, agradezco a las numerosas personas que con sus comentarios y observaciones han hecho posible esta nueva versión del libro. Muy particularmente quiero expresar mi reconocimiento a Gustavo Fernández, Joanne Moller, Gloria Patricia Sierra, Serafín Mercado, Mario Sánchez, Javier Urbina, Javier Gómez, Rebecca Braverman, Silvia Vigil, Nahum Martínez y Arturo Aguilar.

LUIS CASTRO

índice general

1	Introducción	11
2	Consideraciones fundamentales	19
2.1.	Información experimental	19
2.2.	Confiabilidad y generalidad	21
2.3.	Aleatorización	22
2.4.	Estrategias de investigación	22
2.5.	Validez interna y validez externa	25
2.6.	El lugar de la variabilidad	26
2.7.	Replicabilidad	29
2.8.	Control experimental	30
3	Diseños preexperimentales	33
3.1.	Diseños sin nivel de comparación	35
3.2.	Diseño de comparación estática	42
3.3.	Algunos diseños cuasiexperimentales	43
3.4.	Diseños apareados-correlacionales	49
4	Conceptos básicos del diseño experimental	53
4.1.	Experimentos con grupos de sujetos y experimentos con organismos individuales	54
4.2.	El concepto de grupo control	56
4.3.	El concepto de línea base conductual. Un sujeto como su propio control	58
4.4.	¿Un sujeto medido n veces $\equiv n$ sujetos medidos una vez?	62

4.5.	$N = 1$	63
4.6.	Comparación entre línea base y grupo control	65
4.7.	Modelo fijo y modelo aleatorio en el diseño de grupos	68
5	Diseños experimentales. Generalidades y clasificación	71
5.1.	Clasificación de acuerdo con la(s) variable(s) independiente(s)	77
5.2.	Clasificación por tipo de metodología. Diseños conductuales y diseños tradicionales	78
6	Diseños univariantes bicondicionales	81
6.1.	Diseños A-B	81
6.2.	Diseño balanceado simple	92
6.3.	Diseño de apareamiento ayuntado (acoplado)	94
6.4.	Diseño de dos grupos aleatorios	95
6.5.	Diseño de apareamiento directo	98
7	Diseños univariantes multicondicionales	103
7.1.	Diseño reversible A-B-A	103
7.2.	Diseño reversible A-B-A-B	110
7.3.	Diseño reversible A-B múltiple (A-B-A-B)	117
7.4.	Diseños balanceados conductuales	119
7.5.	Diseño de línea base múltiple	120
7.6.	Diseño de más de dos grupos aleatorios	122
7.7.	Diseños contrabalanceados	124
8	Diseños multivariantes	127
8.1.	Diseños reversibles multivariantes	128
8.2.	Diseños factoriales	133
9	Otros diseños	143
9.1.	Diseño reversible con una variable control	143
9.2.	Diseño de criterio cambiante	146
9.3.	Diseños de medidas repetidas dentro del mismo sujeto	146
9.4.	Diseños entre-dentro	149
9.5.	Diseños jerárquicos (grupos dentro de tratamientos)	150

9.6. Diseños incompletos	152
9.7. Diseños de bloques aleatorios	158
10 Evaluación y comparación de los diseños experimentales	161
10.1. Evaluación de los diseños tradicionales	161
10.2. Evaluación de los diseños conductuales	174
10.3. Comparación entre los diseños experimentales tradicionales y los diseños conductuales	185
11 Compatibilidad entre los diseños conductuales y los diseños tradicionales	199
11.1. Metodología	199
11.2. Estructura	201
11.3. Análisis y evaluación de los datos en las ciencias de la conducta	211
11.4. La estabilidad en los datos conductuales. Cuándo y cómo suspender una línea base en investigación con sujetos humanos	219
Referencias bibliográficas	233

1

introducción

ESTE LIBRO SE ha escrito para los estudiosos de la ciencia de la conducta en particular y para todos aquellos que estén interesados en diseñar experimentos en los cuales participen organismos comportantes* en general. El énfasis que se ha hecho en la conducta humana a través de la exposición de los diversos temas se sigue de la suposición básica de que los organismos comportantes son sujetos humanos; sin embargo, muchos de los ejemplos citados fueron tomados de las investigaciones llevadas a cabo con animales. En esta obra, como su título pudiese sugerir, se hace hincapié en la estructura y la lógica de los diferentes diseños experimentales, más que en el análisis estadístico de los datos producidos por un experimento o en los modelos estadísticos que rigen a los análisis o que influyen a los propios diseños.

Aun cuando se ha intentado escribir este libro en términos generales y evitando el empleo de tecnicismos sin dar por lo menos una breve explicación, se espera que el lector esté familiarizado con conceptos experimentales básicos. El haber leído por lo menos una obra introductoria de psicología experimental ayudará considerablemente para la ubicación y comprensión del contenido de este volumen. Dando esto por sentado, se hará referencia a términos como variable independiente, variable dependiente, relación funcional, control experimental, etc., con relativa libertad.

En una parte de esta obra se incluye la presentación y el análisis de diseños preexperimentales. El presentar dichos diseños tiene un doble propósito: *a*) por una parte, estos diseños sirven como punto de comparación para llevar a cabo la evaluación de los verdaderos diseños experimentales. Resulta útil estar familiarizado con lo que no es un experimento, a fin de poder reconocer lo que sí es un verdadero experimento; y *b*) una gran cantidad de "datos" que aparecen en algunos volúmenes sobre educación, sociología y algunas ramas precientíficas

* Por *comportante* se entiende el organismo que emite respuestas en la situación experimental.

de la psicología (como la psicología clínica o las "teorías" de la personalidad) se han derivado de situaciones preexperimentales. De tal suerte, el conocimiento de las limitaciones y los errores producidos por las aproximaciones precientíficas a los problemas del comportamiento permiten colocar a los hallazgos preexperimentales y a los experimentales en su lugar adecuado.

También se discute extensamente la posición inductiva extrema propugnada por Skinner y otros analistas experimentales de la conducta, en la cual se considera innecesario, o de poco valor, el diseñar experimentos. Skinner (Evans, 1968), se precia de reportar que ninguno de sus estudiantes jamás ha diseñado un experimento. Esta posición, sumamente respetable, es el resultado de un tipo de tácticas de investigación que han producido una masa de datos que han dado gran auge a la psicología en los últimos años; como tal, dicha posición merece una seria consideración.

La escuela de Skinner se conoce generalmente como el *análisis experimental de la conducta* (AEC) a nivel básico, y como el *análisis conductual aplicado* (ACA) a nivel aplicado. Resulta paradójico que en la literatura de ambas ramas del análisis conductual experimental aparezcan referencias al diseño experimental, especialmente en la del ACA. El hecho de que el diseño experimental sea tratado y desarrollado por algunos analistas experimentales de la conducta es una justificación, entre otras, para que una parte fundamental de este libro esté dedicada a esta escuela y a sus diseños. Otra justificación es la falta de un libro que trate exclusivamente los diseños experimentales conductuales. Una más, es que su conocimiento permite una mejor evaluación del diseño tradicional.

Lo que aquí denominamos *diseño experimental tradicional* fue por mucho tiempo lo único que se reconocía como diseño experimental. En la mayoría de las obras sobre diseño experimental en el campo de la psicología, se desarrolla exclusivamente este tipo de diseño, su empleo es extenso todavía y no hay indicios de que se le vaya a abandonar. Es por ello que también se le dedica otra parte fundamental de la obra. Esta clase de diseño se encuentra ligado a lo que se ha denominado *método hipotético-deductivo*. Su lógica y estructura serán estudiadas con cierto detenimiento, y sus méritos y deficiencias serán puestos a consideración, al igual que los de los diseños del AEC.

El lector entrenado en el AEC en general, o la metodología del condicionamiento operante en particular, quizá encontrará que es excesivo el número de secciones dedicadas al diseño experimental tradicional. Si esto es así, sería conveniente recordarle que hace apenas algunos años esta cuestión dominaba la literatura en el campo del

diseño experimental. Además, resultaría muy difícil hacer una justa evaluación de los propios diseños del AEC sin tener como referencia los diseños tradicionales. Por otra parte, el lector entrenado en la escuela tradicional, probablemente encontrará excesivo el número de trabajos "conductistas" estudiados. Si hay alguna desproporción en este sentido, ésta será probablemente una consecuencia del ímpetu con que tal aproximación está probando su utilidad. Comoquiera que sea, el lector tendrá oportunidad de enterarse de "la otra versión" y de hacer las comparaciones pertinentes. Incluso, la tesis de este trabajo afirma que, pese a la irreconciliabilidad entre aproximaciones metodológicas tan dispares, es posible hacer combinaciones extremadamente útiles entre diseños de ambas clases. En los últimos capítulos de esta obra se tratará este problema.

Cuando un lector se encuentra frente a un libro con un título como "Diseño experimental" o alguna variante del mismo, pudiera pensar que va a leer una obra que le enseñará cómo planear y ejecutar diversos experimentos; sin embargo, pronto descubre que en el libro se hace más énfasis sobre el primer aspecto que sobre el segundo. Con algunas excepciones, encontrará mayor espacio dedicado a consideraciones acerca de lo que es válido para planear un experimento, acerca de los modelos que corresponden a ciertos diseños y a procedimientos computacionales; poco se dice acerca de la implementación del experimento. Pocas obras (Ferster y Skinner 1957, Sidowsky 1966, Honig 1966) tratan extensamente los aspectos de instrumentación y del manejo de los organismos estudiados. En muchos casos se tiene que recurrir a manuales técnicos o a la antigua práctica de "echar a perder con tal de aprender", a fin de poder contestar preguntas concernientes a la ejecución de experimentos. En resumen, las obras sobre diseño experimental generalmente no enseñan cómo ejecutar experimentos, y ésta no será la excepción. Sin embargo, el lector podrá notar dos diferencias importantes entre este trabajo y el libro típico sobre diseño experimental: por una parte, verá que los diseños estudiados pueden aplicarse a problemas conductuales con un mínimo de restricciones de origen estadístico, y por otra, hallará una colección de diseños con una orientación conductual que no aparece en otros textos sobre este tópico.

Para poder diseñar un experimento que produzca resultados relevantes a la contestación de una pregunta de investigación, el experimentador deberá ser capaz de plantear ésta en una forma específica. La pregunta deberá indicar claramente cuáles son las variables implicadas y qué relación entre ellas se está buscando. La relación generalmente será del tipo: "¿Cuál es el efecto de la variable x sobre la variable y ?" y no más específica. Comoquiera que sea, el planteamiento debe ser

lo suficientemente preciso para poder implementar la pregunta en el laboratorio y tratar de contestarla parcial o totalmente. La pregunta en un principio puede ser demasiado amplia o sin sentido (por ejemplo, ¿por qué se comportan los organismos?); pero una vez que se le ha considerado con seriedad, se le puede reducir a términos que la hagan susceptible de ser tratada por medio de la metodología experimental. El diseño de un experimento requiere, entonces, la reducción de la pregunta de investigación a términos manejables dentro de un laboratorio. La pregunta ¿por qué se comportan los organismos? se podría reducir a la más modesta ¿cómo se comportan los organismos?; ésta, a su vez, se puede especificar aún más: ¿cómo se comporta esta clase de organismos en una unidad de condicionamiento operante bajo determinadas condiciones de privación de alimento? Sucesivamente se podría llegar a una pregunta tan específica como: ¿cuál es la distribución de los tiempos interrespuesta de este organismo bajo una privación de alimento de 20 horas y de acuerdo a un programa de razón fija de 16?

Asimismo, al diseñar un experimento se deben hacer consideraciones acerca de las variables independientes (VI), de las variables dependientes (VD) y de las variables extrañas* (VE) implicadas; además, deben contestarse las siguientes interrogantes: ¿qué variable(s) se va(n) a manipular experimentalmente?, ¿cómo se va a llevar a cabo la medición de los cambios en la(s) conducta(s) de interés?, ¿cómo se van a controlar otras variables que pudiesen afectar a la VD? Las consideraciones primordiales incluyen, además, el rango de la(s) VI que se planea manipular, la forma en que se habrán de cuantificar tanto las VI como las VD, la confiabilidad de los instrumentos de medición o de los registradores humanos, y la correspondencia entre la medida tomada y la definición de la conducta de interés.

El investigador deberá, por supuesto, estudiar minuciosamente la selección del diseño experimental. Esta selección dependerá básicamente, de la forma en que se plantee la pregunta de investigación, de la metodología indicada por el problema o preferida por el experimentador, de la calidad de las instalaciones disponibles, y de los recursos económicos asignados al experimento. Estos últimos aspectos pueden limitar seriamente las opciones del investigador al seleccionar el diseño.

La selección de un diseño experimental, fuera de algunas de las consideraciones prácticas antes mencionadas, depende fundamentalmente de las formas en que se puede circunscribir la pregunta de investigación. Siempre habrá un número de formas alternativas de traducir

* El término *variables extrañas* se refiere a todas aquellas variables ajenas al control del experimentador aunque éste tenga conocimiento de ellas. Se mantiene el término para ser consistente con otras obras en la materia.

la pregunta a condiciones experimentales específicas; de éstas, se debería elegir la más ingeniosa, económica e informativa. En ocasiones, un texto sobre diseño experimental puede servir como "recetario" o "directorío", el cual le indicará al lector cuáles diseños se encuentran a su disposición. Generalmente, si el lector busca con detenimiento, encontrará un diseño acorde con su problema; sin embargo, habrá ocasiones en que, por más que busque en libros y revistas especializadas, no podrá encontrar un diseño apropiado a su pregunta. Es en este caso cuando se producen las contingencias para que él desarrolle un nuevo diseño. Muchos de los diseños con los que ahora estamos familiarizados surgieron de necesidades similares. Una práctica útil que se sugiere al lector que no tiene experiencia en el diseño y ejecución de experimentos, es la de tratar de ponerse en la situación hipotética de planear un experimento y tener la necesidad de "inventar" un diseño. Uno de los resultados más frecuentes es el del "descubrimiento" de diseños que han estado en uso durante mucho tiempo. Cuando esto sucede, el "autor" del diseño debería verse reforzado por haber seguido su propio curso hasta encontrarse con el de investigadores más avanzados, y no debiera sentirse decepcionado por no haber sido el primero en llegar.

Otro aspecto importante del diseño experimental es el análisis que se hace de los resultados de un experimento. En ocasiones, los resultados simplemente se presentan por medio de tablas y gráficas. Si los datos sólo describen el comportamiento de las VI y las VD en cada uno de los sujetos o grupos de sujetos, entonces no es necesario hacer ningún tratamiento elaborado de los datos; sin embargo, cuando la cantidad de datos es demasiado grande, se recurre al uso de algunas idealizaciones estadísticas, como las medidas de tendencia central ("promedios" como la media aritmética o la mediana), las medidas de variabilidad (por ejemplo, la varianza), o algunas otras.

Los datos resumidos de esa manera resultan útiles para comparaciones más complejas, las cuales, a su vez, aportan las bases para la inferencia estadística. Cuando el nivel de análisis de los datos excede al de la simple descripción, se está haciendo inferencia estadística. Algunas prácticas adoptadas en el diseño experimental tradicional se siguen como consecuencia de las restricciones indicadas por los modelos estadísticos relacionados a determinado diseño; por ejemplo, con el fin de obtener distribuciones de datos que luzcan regulares o "suaves" (suposición de normalidad), se requiere que un gran número de datos aparezcan en la distribución. A veces, cuando sólo se obtiene una observación por sujeto, las propias características del modelo estadístico exigen que se tenga una gran cantidad de sujetos en cada condición. Además, las consideraciones estadísticas indican la inclusión

de cierta clase de hipótesis a ser evaluadas. Estas hipótesis son conocidas como hipótesis nulas, y su evaluación constituye uno de los primeros pasos en el proceso de inferencia experimental tradicional. Como el propósito fundamental de la inclusión de las hipótesis nulas es el de aportar indirectamente alguna evidencia empírica en favor de una hipótesis alternativa, este procedimiento contribuye a moldear la conducta del experimentador durante la conducción y evaluación de sus experimentos.

Otra alternativa disponible para un experimentador es la de promediar sus datos y así obtener "curvas de aprendizaje" o alguna otra forma de representación matemática. El trabajo de Hull (1950) es representativo de esta tendencia. Skinner (Evans 1968) argumenta que el promediar los datos para producir ciertas idealizaciones matemáticas es una forma de "puerta falsa" para deshacerse del problema. Sugiere que cuando se encuentren irregularidades o "impurezas" en los datos, lo que se debe hacer es apegarse más al problema y averiguar sistemáticamente en el laboratorio cuáles son las causas de las irregularidades observadas.

Cuando se estudian los diseños tradicionales, generalmente se presenta, junto con el diseño, el modelo estadístico apropiado, sus suposiciones básicas y sus restricciones. También se suele indicar cuál es la hipótesis nula, y cuál es el análisis estadístico que más comprensivamente trata los datos. Para Myers (1972) un texto útil sobre diseño experimental debe "presentar claramente los diseños y análisis de los datos, complementándolos con instancias específicas de aplicación y con ejemplos numéricos" (prefacio de la primera edición, ix). En esta obra se pretende, desde luego, presentar los diseños claramente. También se estudiarán instancias específicas de aplicación; sin embargo, se omitirá el análisis de los datos y el complemento de los ejemplos numéricos. Aún así, se pretende que este libro sea de gran utilidad.

Sería interesante hacer hincapié en que muchos de los diseños experimentales tradicionales se desarrollaron con el fin de resolver problemas de la agricultura o de la biología, y que la psicología posteriormente los adoptó y trató de ajustarlos a sus propios propósitos. Entonces, no debe extrañar que las suposiciones y restricciones estadísticas de algunos de estos diseños parezcan raras o difíciles de cumplir cuando se aplican a problemas conductuales. Los nexos entre las consideraciones estadísticas y la lógica experimental de los diseños tradicionales son tan numerosos, que muchas veces es imposible hacer una diferenciación de ellas. Asimismo, debe extrañar el hecho de que, frecuentemente, un libro sobre diseño experimental comience tratando problemas experimentales y termine tratando problemas puramente estadísticos.

Las objeciones principales a esta estructura estadísticamente orientada de los diseños tradicionales parecen ser legítimas y explican la tendencia de algunos investigadores a referirse a ellos como "diseños estadísticos". Por otra parte, esta denominación no es totalmente justa, ni siquiera objetivamente descriptiva. Los diseños tradicionales se pueden considerar como experimentales, puesto que básicamente representan "esqueletos" o esquemas de acción para la aplicación de la metodología experimental; el empleo de esta última les da su carácter experimental. Para dilucidar más claramente esta controversia estadístico-experimental de la naturaleza de los diseños, sería conveniente considerar por separado al usuario y a sus instrumentos de trabajo. Los diseños experimentales (tanto los tradicionales como los del AEC) existen independientemente de sus usuarios, y la ausencia, presencia o abuso de los análisis cuantitativos (por ejemplo, los estadísticos) de los datos que se obtienen mediante su utilización se determina por el usuario mismo. Los esfuerzos de esta obra estarán dirigidos a separar, hasta donde sea posible, las consideraciones experimentales y la lógica de los diseños, de las consideraciones estadísticas enfatizando las primeras y minimizando las últimas. Se trata, pues, de un libro sobre el "diseño experimental sin estadística", y el significado de este título se halla claramente representado en el párrafo precedente.

Como, a pesar de todo, la estadística mantiene diversos grados de relevancia para el diseño experimental, y como este libro fue escrito para el lector que no está familiarizado con la estadística, se recomienda al lector interesado en este aspecto el estudio de las obras de los siguientes autores (dichas obras cubren extensa y detalladamente tanto los fundamentos estadísticos, como los procedimientos computacionales necesarios para la comprensión de la estadística asociada al diseño experimental): Cochran y Cox (1971), Edwards (1968), Hays (1963), Kirk (1968, 1972) McNemar (1969), Myers (1972) y Winer (1962), Keppel (1973) y Glass y Stanley (1970).

Finalmente, sintetizando el contenido del presente libro, se puede decir que cubre los siguientes aspectos del diseño experimental en la ciencia de la conducta: fundamentos (capítulos 2, 3 y 4), exposición (capítulos 5, 6, 7, 8 y 9), evaluación y selección (capítulo 10), así como la compatibilidad entre diseños conductuales y tradicionales, y el desarrollo de nuevos diseños (capítulo 11).

2

consideraciones fundamentales

HAY UNA SERIE de consideraciones sumamente relevantes para todo aquel que intenta comprender o participar en la experimentación conductual, y muy particularmente para quien desea diseñar un experimento con la participación de organismos comportantes. Por lo menos, los siguientes aspectos merecen una seria consideración: *a*) información experimental; *b*) confiabilidad y generalidad; *c*) aleatorización; *d*) estrategias de investigación; *e*) validez interna y validez externa; *f*) variabilidad, replicabilidad, y *g*) control experimental. A continuación se presenta un breve estudio de cada una de ellas.

2.1. INFORMACIÓN EXPERIMENTAL

Campbell (1957) afirma que para obtener un mínimo de información científica útil, se requiere obtener datos de dos observaciones minuciosas, con el fin de poder hacer por lo menos una comparación formal. La experimentación en general, y el diseño experimental en particular, implican comparaciones obtenidas bajo condiciones controladas. Planear situaciones en las cuales se pueda lograr al menos esas dos observaciones o medidas es una de las funciones características del diseño experimen-

tal; sin embargo, cumplir dicho requisito no garantiza que un diseño particular sea un verdadero diseño experimental. El hecho de que las observaciones sean realizadas sobre el mismo individuo (o grupo de individuos) o sobre diferentes individuos (o grupos de individuos) tampoco determina la calidad experimental del diseño o de los datos que la situación arroja.

Lo mismo sería válido para la precisión o extensión de las observaciones. Una investigación puede producir observaciones sumamente detalladas, sistemáticas y extensas, y aun así no ser de tipo experimental. Lo que determina la categoría experimental de diseños e información es la aplicación de los procedimientos de la metodología experimental. Entre éstos, el máspreciado es el control experimental sobre las variables que pueden determinar el comportamiento de las *variables dependientes* (VD), a saber: las *variables independientes* (VI) y las *variables extrañas* (VE). Cuando el control experimental no se puede aplicar para separar los efectos de las VE de los de las VI sobre las VD, entonces se puede asegurar que los diseños con los que se está trabajando son preexperimentales y que la información así obtenida no es experimental.

Hay por lo menos dos criterios generales para diferenciar la información experimental de otros tipos de información. El primero está relacionado con las operaciones que se llevan a cabo para su obtención, entre ellas. *a*) la organización de las condiciones para contestar preguntas específicas de investigación (presencia de alguna forma de diseño experimental); *b*) la manipulación de VI de acuerdo a un plan predeterminado (aunque sea vagamente); *c*) la eliminación o control de las VE; *d*) el empleo de técnicas de medición o expresión cuantitativa de los datos, y *e*) como consecuencia de las anteriores, el escenario donde se coloca al organismo es diferente de su ambiente natural y tiende a ser artificial.

El segundo criterio se relaciona con los resultados que se obtienen como consecuencia de las operaciones anteriores: *a*) se obtiene información específica, generalmente relacionada con la interrogante de investigación y limitada por las condiciones del diseño; *b*) es marcadamente diferente de la que se obtendría por medio de la observación sistemática o naturalista, ya que muchas de las respuestas que se obtienen en un laboratorio difícilmente ocurrirían de manera espontánea, y *c*) tiende a ser más conclusiva que la información preexperimental (es *más inequívoca* respecto al establecimiento de relaciones causales). Los aspectos siguientes pueden servir para valorar mejor la información de tipo experimental.

2.2. CONFIABILIDAD Y GENERALIDAD

Como características importantes en la obtención de información experimental tenemos la confiabilidad de los datos y su generalidad. La confiabilidad es la consistencia que muestran los datos a través del tiempo (dadas las mismas condiciones), el grado de acuerdo entre observadores independientes o la consistencia de los datos a través de diferentes investigaciones. Sidman (1960) argumentó que la única forma razonable de obtener datos confiables es por medio de la replicación. La evaluación reciente de los índices de confiabilidad conductuales muestra que éstos se encuentran alterados en una gran cantidad de estudios (Fernández, 1972, comunicación personal). También se ha encontrado (Romanczyk y colaboradores, 1973) que cuando hay un supervisor cuya presencia es obvia, la proporción de acuerdo (confiabilidad) entre los observadores es notablemente más alta (a menudo tanto como 0.40 o más alta) que cuando su presencia se encubre. Estos hallazgos le dan a Sidman la razón en una forma indirecta.

El concepto psicométrico de confiabilidad también es cuestionable (Cañedo, 1971, comunicación personal). Cuando se pretende que las respuestas de una persona o grupo a un test psicológico sean lo más análogas entre sí en dos ocasiones diferentes, se trata de "aminorar" el efecto de las interacciones que la persona ha tenido con el ambiente en el lapso transcurrido entre ambas ocasiones. Este es el caso en que las puntuaciones de los tests serían "confiables" (estarían altamente correlacionadas) y el cual es poco frecuente. Por lo contrario, si el test fuese lo suficientemente sensible como para advertir los cambios en la conducta de la persona, entonces las puntuaciones no serían "confiables" en el sentido tradicional, y el test se suprimiría. Así, la confiabilidad dependerá de que las relaciones funcionales encontradas en un experimento se puedan reproducir consistentemente un cierto número de veces, es decir, que la relación sea "sólida" y no un accidente fortuito.

El mismo Sidman indicó que algunas formas de replicación proporcionan algo más que indicaciones acerca de la confiabilidad. La replicación sistemática, por ejemplo, provee bases sólidas para la generalización. Un experimento deberá replicarse exitosamente en diferentes laboratorios, a fin de que los datos obtenidos sean generalizables. Este autor consideró las siguientes formas de generalidad: *a*) de sujeto; *b*) interespecie; *c*) de variables; *d*) de proceso, y *e*) metodológica. Criticó, además, la aproximación tradicional al estudio de este proble-

ma y ofreció una alternativa más analítica basada en el paradigma intersujeto.

2.3. ALEATORIZACIÓN

Cuando se trata de comparar los datos de dos o más grupos de sujetos distintos, hay una operación que puede diferenciar a los diseños experimentales de los preexperimentales: la aleatorización. Si el investigador se ve forzado a utilizar grupos formados siguiendo criterios educativos, sociales, etc., el valor de las afirmaciones que podrá hacer es muy limitado. Además, no importa en qué forma organice las comparaciones de interés (su "diseño"), ni el número de medidas que tome: sus datos siempre serán preexperimentales. Si el control del experimentador estuvo ausente para la asignación de los sujetos a diferentes condiciones, otras formas más rigurosas de control de VE es probable que también estuviesen ausentes. La aleatorización proporciona al investigador un elemento que permite que las decisiones acerca de la asignación de los sujetos a los diversos tratamientos no sean el producto de alguna distorsión, regla o parcialidad arbitraria y sistemática. La aleatorización evita que una arbitrariedad seguida sistemáticamente por el investigador o por alguien más, haga tendenciosos los resultados. El azar es arbitrario, pero sin sistema. La suposición fundamental detrás del empleo del azar es que la expectancia de que los efectos de las VE sobre la VD serán distribuidos equitativamente entre las diferentes condiciones se ve maximizada. Esta argumentación se basa en la teoría de la probabilidad, y el lector podrá crear una pequeña situación "experimental" para comprobar la solidez de estos argumentos.

2.4. ESTRATEGIAS DE INVESTIGACIÓN

¿HIPÓTESIS O PREGUNTA DE INVESTIGACIÓN?

Una hipótesis es una afirmación. Esta afirmación generalmente indica una relación específica entre dos o más variables. Algunos autores, como McGuigan (1971), consideran que una hipótesis es una solución tentativa de un problema. Otros, como Skinner (1938) o Sidman (1960), la conceptúan como una forma de adivinanza más o menos inútil. El hecho es que las hipótesis, como se emplean predominantemente en el campo de la psicología, especifican un resultado como consecuencia de un argumento antecedente. Un problema fundamental en el empleo de esta forma de las hipótesis, forma condicional, lo constituye

la solidez del argumento antecedente. Este argumento pudiera haber sido derivado de datos experimentales, pero también pudiera condensar la forma de pensar, las observaciones o las corazonadas de un investigador. Aunque las diferentes formas de pensar, observaciones o corazonadas de diversos investigadores puedan ser respetadas, esto no justifica que formulen hipótesis que, a largo plazo, implicaran el esfuerzo desperdiciado de un gran número de personas (incluyendo a los lectores de los resultados de la investigación).

Tradicionalmente, en otras ciencias, una hipótesis no se formula sino hasta que se ha acumulado una gran cantidad de información experimental (o de observación sistemática en ciencias como la astronomía). No sucede lo mismo en el campo de la psicología. Parece ser que la afirmación de Murray (en Plutchik, 1968), en el sentido de que en la psicología se tiende a saltar etapas del proceso histórico-científico, se ve confirmada por esta disposición a la formulación de hipótesis sin la evidencia empírica suficiente. Estas consideraciones acerca de la formulación de hipótesis en la psicología es relevante, porque la elección de un diseño experimental se ve fuertemente influida por aquélla.

Un tipo particular de hipótesis que se plantea en innumerables ocasiones es lo que se conoce como *hipótesis nula*. Este tipo de hipótesis desempeña un papel fundamental en el proceso *estadístico* de análisis y evaluación de datos. La suposición básica en este caso es la de no-diferencia y su planteamiento debe ser exacto, sin ambigüedades (Fisher, 1966). La forma general de una hipótesis nula puede ser como sigue: "Si la VI no tiene efecto sobre la VD, entonces no se espera que haya diferencias en los valores de la VD en las distintas condiciones experimentales."

Una vez que se ha planteado alguna forma particular de esta hipótesis, se elige un diseño experimental que permita su evaluación. Se efectúa el experimento y se recogen los datos. Éstos se procesan y se disponen de tal forma que puedan ser incluidos en las fórmulas estadísticas indicadas por el tipo de diseño y por la clase de medidas tomadas. Los números que se obtienen (por ejemplo, las diferencias entre dos "promedios") se comparan con otros números derivados teóricamente y se toma una decisión. Esta decisión es probabilística.

En general, se considera que una gran diferencia es poco probable y que una pequeña diferencia es muy probable. Al respecto, las reglas para determinar lo que es poco probable (0.05, 0.01, 0.001, 0.0001) o muy probable (0.50, 0.90, 0.95, 0.99, 0.999, 0.9999) son arbitrarias y han permanecido como algo "incuestionablemente correcto" (véase Skipper y colaboradores, 1967). Cuando un evento (como una diferencia entre dos "promedios") es poco probable, entonces se considera

que la influencia del azar no es tan grande como para producirlo. Si se asume que el control experimental fue estricto, entonces se atribuye el evento (la diferencia) a la VI y se rechaza la hipótesis nula. Por otra parte, si el evento es muy probable, entonces se considera que el efecto de la VI no fue lo suficientemente fuerte como para superar a lo que el azar (o sea, la simple distribución esperada del comportamiento desordenado de cualquier serie de números) hubiese producido, y entonces se acepta la hipótesis nula.

Fernández (1969) equipara el término azar con los vocablos "ignorancia" o "variabilidad de muestreo", intentando con ambas formas indicar "falta de control experimental". Al respecto, afirma: "El rechazo de la hipótesis nula sólo autoriza a concluir que el «azar» no es la «causa» de lo observado. Para atribuir causalidad a alguna variable, es preciso haberla observado en condiciones fijas, controladas."

Otro grave problema en el planteamiento de hipótesis (nulas o de otro tipo) consiste en que la afirmación que se hace constituye una sola posibilidad entre muchas. Un experimento puede haber sido precedido por experimentos similares, y éstos pueden haber determinado o sugerido la forma de la hipótesis. El experimentador puede tener bastante certeza acerca de lo plausible que es su hipótesis; aun así, nada garantiza que los datos la van a apoyar. Quizá hay n alternativas igualmente probables y tal vez la menos esperada de todas sea la que ocurra. Sidman (1960) señaló: "Cuando un investigador lleva a cabo un experimento sin tratar de probar una hipótesis, su vida está llena de sorpresas."

Finalmente, se puede entender que cuando el investigador no está comprometido con ninguna hipótesis, trabaje "con los ojos más abiertos". Como su interés no consistirá en seguir paso a paso la forma en que la hipótesis va siendo sometida a prueba, es probable que se dedique a registrar y reportar cualquier regularidad que considere importante. Alguna relación funcional relevante podría estar operando sin tener en cuenta alguna de sus observaciones, a las que les ha prestado el tiempo y la atención que, de otra manera, hubiera dedicado a considerar la hipótesis.

Una alternativa a la estrategia de investigación que se centra en la formulación de una hipótesis es la estrategia que parte de una pregunta de investigación. Cuando el experimentador formula una pregunta de investigación, no se ve forzado a "adivinar" las formas en que se relacionarán sus variables. No dirige sus esfuerzos a la búsqueda de una relación que se postuló de antemano como "solución tentativa". El tipo de pregunta que generalmente se plantea de acuerdo a esta estrategia (método inductivo) sigue una forma semejante a: "¿Qué pasará en

la VD cuando manipule la VI en tal forma?" La pregunta puede referirse a un asunto extremadamente específico, y su contestación podría aportar poco o nada a la formulación de una teoría. Esto podrá ser cierto para un gran número de experimentos de esta clase cuando se les considera separadamente. Con el tiempo, los datos que estos experimentos han producido podrán agruparse y organizarse para establecer uniformidades o "leyes" que pudieran contribuir sustancialmente a la formación de una teoría. Parece ser que el momento en que una verdadera teoría, en el sentido en que se le considera en otras ciencias, sea aceptada en el campo de la psicología, todavía está lejano. La ciencia de la conducta (psicología) se encuentra en el estadio primitivo de la descripción y análisis de los datos. De este análisis ya han surgido algunas uniformidades o leyes y de ésta, algunas son claramente relaciones funcionales; por ejemplo, ya se sabe con certeza que cada programa de reforzamiento positivo produce formas de conducta impresionantemente regulares en organismos de diferentes especies, incluyendo al hombre. Desafortunadamente, los logros de este tipo sólo constituyen un pequeño grupo, y aún se espera su formulación formal como "leyes". A este respecto, algunos lectores podrían argüir que al consultar un diccionario de psicología encontraron docenas de "leyes". Esto es cierto, pero merece comentarios aclaratorios. Primero, algunas de las citadas "leyes" están tan débilmente formuladas y documentadas de una manera tan pobre, que ni siquiera se les puede considerar como reglas válidas. Segundo, algunas otras, como la conocida ley de Weber-Fechner, aun cuando se encuentren sólidamente formuladas y documentadas en forma abundante, carecen de generalidad. Cuando el rango de la VI en el cual la "ley" es válida se encuentra severamente restringido (como parece ser el caso de la ley Weber-Fechner), entonces su valor como ley se ve disminuido. Estaríamos tratando, entonces, con una "ley" que no siempre se cumple (véase Stevens 1951; D'Amato 1970).

2.5. VALIDEZ INTERNA Y VALIDEZ EXTERNA

Campbell y sus colaboradores (Campbell y Stanley, 1966; Campbell, 1967; Stanley, 1967) han planteado la división entre dos fuentes de validez (interna y externa) y han desarrollado formas para evaluar los diseños experimentales y preexperimentales en este respecto.

El control que se ejerce sobre las VE en los verdaderos diseños experimentales lo relacionan con la validez interna. Algunas de las fuentes que operan contra dicho control son la maduración, la historia del sujeto, la "mortalidad" (pérdida de sujetos a través de un experimento)

y la selección. Estos aspectos se tratarán detalladamente en el siguiente capítulo.

Los aspectos que se relacionan con la representatividad o generalización estadística constituyen la fuente de validez externa. Uno de estos aspectos se relaciona con la obtención de medidas reactivas, o sea, aquellas que producen una modificación al medir un cambio (Campbell, 1967). Es evidente que muchas de las medidas que se obtienen tanto a nivel experimental como preexperimental en las ciencias del comportamiento son de tipo reactivo. En muchas de las situaciones preexperimentales, donde se le aplica al sujeto un test o cuestionario, se le hace reaccionar a la estimulación contenida en el "instrumento de medición".

Otro aspecto de la validez externa se refiere a los efectos reactivos de la propia situación experimental. Campbell y Stanley han analizado este aspecto, básicamente en relación a la participación de sujetos humanos en experimentos. Estos investigadores señalan que sería cuestionable extrapolar los hallazgos de laboratorio a escenarios no experimentales; sin embargo, si se amplía este concepto de los efectos reactivos de la situación experimental a una gran variedad de organismos, entonces se verá que este problema se relaciona también con la validez interna, ya que se considera que la producción de conductas emocionales ante la situación de laboratorio es una VE que debe ser controlada. Incluso, tanto en el laboratorio como en escenarios "naturales" se han desarrollado técnicas para controlar esta VE (por ejemplo, la inclusión de un período de adaptación antes de la fase de observación o registro).

Los otros aspectos de validez externa enfatizados por Campbell y Stanley son: a) la interacción entre parcialidad en la selección de sujetos y la VI, y b) la interferencia entre los tratamientos múltiples. Esto se refiere al efecto residual que ha producido un tratamiento cuando el mismo sujeto es sometido a múltiples tratamientos experimentales. Este último aspecto se trata con mayor detalle en otro capítulo.

2.6. EL LUGAR DE LA VARIABILIDAD

Hasta hace pocos años el papel que desempeñaba la variabilidad como parte fundamental en el análisis y evaluación de los datos psicológicos y sociales era incuestionable. El trabajo de Sidman (1960) ha sido determinante para cambiar esta situación. Para este investigador, la variabilidad puede ser identificada experimentalmente y, también, puede caer bajo el control del experimentador. Sidman aboga por el abandono de la noción de que la variabilidad constituye el objeto

de estudio de la psicología y de disciplinas afines. La noción animista un tanto vaga, de que la variabilidad es algo intrínseco a la conducta de los organismos, la cuestiona considerablemente y le otorga el beneficio de la duda provisionalmente. Afirma Sidman que, antes de dilucidar la cuestión de la variabilidad intrínseca, se requiere el desarrollo de las técnicas de control y de medición con los eventos conductuales hasta su máximo. Como dicho límite parece estar muy lejos de nuestro alcance, la cuestión permanece sin resolverse.

Sin embargo, Sidman ofrece una alternativa a la variabilidad intrínseca. Su razonamiento se podría plantear así: "No toda la variabilidad que observamos en el comportamiento de los organismos es intrínseca. Por lo contrario, una gran porción de la variabilidad del comportamiento es una función de la interacción del organismo con las variables ambientales. Como un buen número de estas interacciones ya se han estudiado y puesto bajo el control del experimentador, entonces una (gran) parte de la variabilidad observada puede ponerse bajo control experimental o imponerse a voluntad. La alternativa principal es la de tratar las variaciones como ejemplos de orden, más que de capricho, en la naturaleza" (pág. 142).

La división que Sidman hace de la variabilidad, entonces, es: por un lado se halla la variabilidad intrínseca, la cual es cuestionable y aún no se ha probado su existencia o no existencia; por otro lado se encuentra la variabilidad impuesta, la cual es una función de las operaciones de control experimental. Cuando la variabilidad es impuesta, entonces se le puede explicar por medio de demostraciones, en las cuales se le imponga o se le reduzca a voluntad.

Sidman señala que los conceptos de variabilidad y línea base son mutuamente excluyentes. Una de las características fundamentales de la línea base (registro preciso y sistemático) de un tipo específico de respuesta del organismo es la *estabilidad*. Una línea base, para ser útil, debe ser estable; o sea, no debe variar arbitrariamente. Un experimentador, antes de efectuar alguna manipulación de su VI, deberá esperar hasta que su línea base de interés se estabilice. Una línea base que contenga pocas observaciones puede no representar adecuadamente el nivel habitual de la respuesta de interés. Esta línea base puede tener variaciones o desviaciones repentinas alrededor de un valor central. Entonces, la finalidad del científico que trabaja con la tecnología de la línea base (la de AEC) consiste en reducir la variabilidad hasta donde le sea posible. Por tanto, este uso "de la variabilidad" está en conflicto con su empleo tradicional.

Una de las técnicas estadísticas más eficaces para el análisis de datos de un gran número de diseños experimentales es el análisis de va

rianza (AVAR). Este tipo de análisis se puede utilizar para evaluar los datos de un diseño en el cual se mide repetidamente a los mismos sujetos, de un diseño en que se manipulan dos o más VI, y de muchos más.

Este análisis, como su nombre lo indica, está basado en la variabilidad de los datos (varianza); por ejemplo, cuando se comparan los efectos de varios tratamientos (valores de una VI) sobre la VD, se utiliza un diseño de más de dos grupos al azar. La suposición estadística detrás del AVAR (análisis de varianza), en este caso particular, es que si la hipótesis nula resulta cierta, la varianza producida por los diversos tratamientos debería ser aproximadamente igual a la varianza (o variabilidad) de los sujetos; es decir, la variabilidad debida a diferencias individuales no controladas experimentalmente debiera equipararse con la variabilidad producida al agrupar aleatoriamente a los sujetos bajo diferentes grupos, a los cuales les fueran dados tratamientos inefectivos. Ahora bien, factores como el número de sujetos asignados a cada condición pudieran afectar considerablemente a la relación entre ambas varianzas; así, a mayor número de sujetos, menor será la variabilidad de los mismos. Entonces, si la variabilidad de los tratamientos (la cual se va a dividir entre la variabilidad de los sujetos) permanece más o menos constante, mientras la variabilidad de los sujetos se reduce a medida que su número aumenta, se puede llegar a obtener una estadística suficientemente grande como para rechazar la hipótesis nula y *atribuir los efectos observados a los tratamientos* (véase Edwards 1972). Nótese que todo lo que se considera en estos casos son las relaciones entre entidades estadísticas, y no entre la VI y la VD como debiera ser.

Esto no significa que el AVAR se debiera desechar. Lo que quiere decir es que, a pesar de su poder, este análisis es vulnerable a los efectos estadísticos de sus relaciones numéricas. Estos efectos estadísticos se pueden confundir con los verdaderos efectos experimentales, y se debe considerar que hay "trucos" estadísticos para hacer que la variabilidad opere en una forma espuria, tendiendo a "probar" algo que no se ha debido a un verdadero efecto experimental. Junto con esta falla del AVAR, tenemos su insensibilidad para advertir verdaderos efectos experimentales. Un ejemplo de esto es el caso en que un verdadero efecto de la VI sobre la VD no es advertido por el AVAR debido al escaso número de sujetos en cada condición. Cuando hay pocos sujetos en una condición experimental (o grupo de sujetos), cualquier puntuación de la VD que se desvíe de la media aritmética (el promedio) contribuirá considerablemente a aumentar la varianza "de los sujetos". El resultado de unas cuantas puntuaciones desviadas

es el de producir un número que apoye a la hipótesis nula, indicando, por tanto, que el tratamiento experimental no es efectivo.

Esta crítica, que en este caso se aplica a la técnica estadística del AVAR, podría ampliarse sin mucha dificultad a otras técnicas estadísticas. Un colaborador del autor afirmó recientemente: "Cuando se emplea un diseño experimental tradicional, prácticamente cualquier cosa se puede «probar», siempre y cuando se tenga un número suficientemente grande de sujetos." Esta afirmación, expresada en un lenguaje más técnico, sería: "Cuando se aplica una técnica de análisis estadístico a los datos de un diseño experimental tradicional, a medida que el número de sujetos aumenta (llenando así las suposiciones del análisis), disminuye la eficacia de la técnica para separar los efectos experimentales de los «seudoeftectos» estadísticos."

Recuérdese que la variabilidad es la "materia prima" de las técnicas estadísticas como el AVAR. Recuérdese también que estas técnicas estadísticas determinan muchas de las características de los diseños experimentales tradicionales y, además, muchas de las prácticas dominantes en la experimentación en psicología durante los últimos cuarenta años por lo menos. Sidman considera el problema de controlar la variabilidad como crítico para alcanzar la generalidad. "La prueba máxima de la generalidad es la replicación, y las fuentes desconocidas o incontroladas de variabilidad son las únicas barreras potenciales para la replicación" (pág. 151).

Señala, además, que la forma en que el científico aplicado puede tratar el problema es diferente de la del científico puro. Mientras éste se halla en condiciones de tratar de controlar la variabilidad, el científico aplicado tendrá que acostumbrarse a aceptarla como es.

2.7. REPLICABILIDAD

Para poder lograr generalidad, los resultados de un experimento deben ser reproducibles. Si un experimento produjera resultados excepcionalmente raros (singulares), irreproducibles, entonces poco o ninguno sería su valor científico. Si un cambio conductual se produce en un solo sujeto o en una sola ocasión y no es posible repetirlo, puede tratarse de un caso raro digno de atención, o de una combinación extraña de circunstancias que pudiera ser interesante, pero su relevancia para la formulación de hipótesis o leyes sería nula.

Por otra parte, si los resultados de un experimento se confirman repetidamente, bajo diferentes circunstancias, entonces estamos tratan-

do con un fenómeno replicable y estamos estableciendo así su generalidad. Si, a pesar de las diferencias, la relación funcional encontrada originalmente se repite, se trata entonces de un hecho empírico real y no de un artefacto producido por un procedimiento particular. La credibilidad y consistencia de un hallazgo se demuestran a través de diferentes formas de replicabilidad.

De este análisis se deriva una definición de replicabilidad: en términos generales, la *replicabilidad se refiere a la posibilidad de reproducir en una forma clara y consistente los resultados de un experimento*. Esto puede hacerse, esencialmente, bajo condiciones equivalentes o bajo condiciones diferentes. Sidman (1960) diferenció dos clases principales de replicabilidad: *directa* y *sistemática*. La primera es aquella en la cual los datos son reproducidos por el mismo investigador (en el mismo o en diferentes sujetos), esencialmente bajo las mismas condiciones. En la segunda, los resultados se replican bajo condiciones claramente diferentes de las originales; esto puede hacerlo el mismo investigador o un investigador diferente en otro laboratorio.

El propio Sidman señaló que mientras el primer tipo de replicabilidad incrementa considerablemente la confiabilidad de los datos, el segundo produce (además de confiabilidad) generalidad e información adicional (pág. 112). Al llevar a cabo un experimento con diferente procedimiento del utilizado originalmente y aun así producir la misma clase de relación funcional, la generalidad de los resultados se incrementa en forma notable. Dingman (1968) señaló tres procedimientos para constatar la reproductibilidad de los resultados. El primero de ellos consiste en muestrear un grupo de datos (de uno o dos sujetos, o de parte de un tratamiento experimental) de una matriz de datos (el total de los datos) de un experimento. Si el experimento es confiable, entonces la muestra deberá producir resultados consistentes con los del estudio en general. El segundo está relacionado con la reproductibilidad de los resultados con diferentes muestras de sujetos; y el tercero, con la validación cruzada de un hallazgo con diferentes procedimientos y distintos métodos de análisis de los utilizados originalmente.

2.8. CONTROL EXPERIMENTAL

Esta es la característica distintiva de la metodología científica. El grado en que una ciencia puede someter a control experimental a los diversos fenómenos de su estudio indica el nivel de desarrollo que ésta ha alcanzado. A mayor grado de control experimental, la información obtenida es mayor y más confiable. Aunque el término *control expe-*

perimental se emplea en ocasiones para denotar ciertas técnicas que se han utilizado en la evaluación de los cambios producidos experimentalmente (Sidman, 1960; Johnson y Solso, 1971), su principal significado se relaciona con el manejo que un investigador hace de las variables relevantes en su trabajo.

Concretamente, el control experimental se refiere a los siguientes aspectos: *a*) la manipulación de las VI; *b*) la minimización de la participación de las VE, y *c*) el registro de los cambios en la VD que ocurren en función de las manipulaciones de las VI. En la ciencia de la conducta en particular, el control experimental se refiere a la manipulación de algunas de las variables que determinan la conducta. Esta manipulación se puede demostrar, haciendo que la conducta covaríe de acuerdo a cambios específicos en los procedimientos experimentales.

Otra forma de demostrar el control experimental es mediante la imposición de variabilidad en los datos; es decir, por medio del incremento o disminución de la variabilidad de la conducta hasta un nivel predeterminado. Sidman consideró que se puede demostrar un grado mayor de control cuando se lleva a una conducta particular a través de una serie gradual de estados diferentes.

3

diseños preexperimentales

AQUELLAS SITUACIONES DE investigación que de alguna manera se aproximan a la verdadera experimentación se consideran en esta obra como preexperimentales. Más específicamente, hay un gran número de ejemplos sobre investigación educativa, social y psicológica, en los cuales el investigador no puede controlar las variables que afectan al objeto de su estudio. Tales situaciones son consideradas como preexperimentales, ya que, aun cuando tratan de incorporar algunos de los procedimientos experimentales, no aplican el principal de ellos: el control experimental.

Un diseño se considera preexperimental cuando es incapaz de producir por lo menos una comparación formal, como señala Campbell. Hay varias formas de obtener tal comparación formal. Una de ellas es la de comparar un grupo de sujetos, al cual se le aplica un tratamiento experimental (grupo experimental), con otro grupo de sujetos, al cual no se le aplica el tratamiento (grupo control). La comparación se efectúa con los datos obtenidos de una medición u observación en cada uno de los dos grupos. Otra forma de lograr por lo menos una comparación consiste en medir al mismo sujeto o grupo de sujetos antes de la aplicación del tratamiento experimental (VI) y después de la misma. Cuando un diseño no plantea formalmente la inclusión de

un punto o nivel de comparación, obviamente queda descartado como una fuente para obtener información científica. Un diseño con un solo grupo, o un diseño en el cual la única serie de medidas u observaciones son las realizadas después de la aplicación de la VI, no puede ser considerado como experimental.

Sin embargo, el cumplir con la característica de producir por lo menos una comparación formal no garantiza que un diseño sea experimental. Así, habrá diseños que, a pesar de hacer una o más comparaciones formales, no son experimentales; por ejemplo, cuando la manera en que los sujetos fueron asignados a un grupo no dependió del investigador, sino de decisiones más o menos arbitrarias como las de un director escolar, entonces no importa que se haga una observación antes y otra después del tratamiento "experimental"; de todos modos, el diseño no será experimental. Tampoco se puede considerar como experimental un diseño en el cual cada uno de los grupos implicados en la comparación representa a una cultura diferente.

Una forma de aproximación preexperimental que frecuentemente se ha pretendido considerar como experimental es la de la psicología clínica o ramas afines. Aquí, una de las fallas más sobresalientes radica en la identificación, definición y medición de las VD. Supóngase que el objeto de estudio de un investigador es un individuo con "personalidad psicopática". La identificación, definición y "medición" de una o más VD que sean "índices de personalidad psicopática" dependen de criterios subjetivos o, en el mejor de los casos, de las puntuaciones obtenidas en un test, al cual se le atribuye la propiedad de "medir la personalidad psicopática". Un diseño que compara dos series de evaluaciones sucesivas de "mejoría" o "recaída", carece de datos útiles en lo que respecta a las observaciones anteriores al tratamiento; y en lo que respecta a las observaciones posteriores al mismo, es un diseño preexperimental. Cuando la VD (un número obtenido en un test, en nuestro ejemplo) no guarda una relación biunívoca con el evento medido, entonces el diseño puede servir para comparar dos series de observaciones con un comportamiento más o menos arbitrario de las puntuaciones del test y no para evaluar un tratamiento que ocurrió en medio de las dos series.

Cuando el investigador está observando los efectos de una VI "natural" (como el desarrollo biológico) sobre determinada VD, no importa lo ingenioso de su diseño, pues éste no podrá ser experimental. La falta de control, por parte del investigador, sobre la VI descalifica al diseño como experimental. Lo mismo sucede cuando, aprovechando algún escenario natural, el investigador no puede controlar las VE que afectan a su VD. Como se puede apreciar, un diseño, para ser expe-

rimental: *a*) debe permitir al menos una comparación entre dos series de datos; *b*) presupone la intervención del experimentador para la asignación de sujetos a grupos (o de tratamientos a sujetos); *c*) requiere una correspondencia biunívoca entre los valores de la VD y el evento medido (como sucede cada vez que un organismo aumenta un número en un contador); *d*) requiere que el investigador manipule directamente la VI; *e*) requiere que el experimentador tenga el suficiente poder para controlar las VE (ya sea eliminándolas o manteniéndolas constantes), y *f*) no basa sus decisiones en suposiciones estadísticas o de alguna otra especie.

En las secciones siguientes se analizan detalladamente diseños que no cumplen con uno o más de los requisitos planteados anteriormente.

3.1. DISEÑOS SIN NIVEL DE COMPARACIÓN

Es útil comenzar el estudio de diseños específicos a partir de lo que no es un verdadero diseño. La primera característica de los diseños experimentales mencionada anteriormente es la de permitir comparaciones; cuando no hay un nivel de comparación, esto no se puede llevar a cabo. Este es el caso de nuestro primer diseño preexperimental.

3.1.1. Diseño de un solo grupo

Supóngase que un maestro emplea un sistema novedoso de enseñanza. Llega a una escuela secundaria y consigue autorización para probar su método. Inicia sus clases utilizando su nuevo método y al terminar el año escolar, examina a los estudiantes y encuentra que el promedio de calificaciones es notablemente alto. Esto, para él, es una prueba contundente de que su sistema de enseñanza es efectivo. Él puede demostrar: *a*) que todos los sujetos fueron sometidos a su “tratamiento”, y *b*) que inmediatamente después del tratamiento, se sometió a los sujetos a una medición. Como las mediciones fueron altas, él las atribuye a lo que las precedió, o sea, a su método de enseñanza. “La causa de las calificaciones altas —diría— fue mi método.”

No obstante, se puede observar fácilmente que las calificaciones altas se podrían haber obtenido sin la aplicación del nuevo sistema o método de enseñanza. Se podría argüir, por ejemplo, que los alumnos ya lograban altas calificaciones desde antes de que el maestro utilizara el nuevo método. En otras palabras, el no haber tomado una medida inicial del rendimiento de los sujetos en la VD de interés invalida cual-

quier afirmación, relacionando la VI (el método) con la VD (las calificaciones).

Una manera de solucionar el problema de la falta de una medida inicial sería utilizar el diseño de un solo grupo mediante pretests y postests, el cual analizaremos a continuación.

3.1.2. Diseño de un solo grupo con pretests y postests

Cuando se carece de un verdadero control experimental sobre las variables (VI o VE) que afectan a la(s) VD, entonces el tener un nivel de comparación dentro del mismo grupo de sujetos es insuficiente para hacer una evaluación del tratamiento. El presente diseño constituye un ejemplo de esta situación. Es cierto que al medirse a los sujetos antes (pretest) del tratamiento y después de éste (postest), se puede comparar el cambio en algún índice de la conducta de los sujetos. En ocasiones, el cambio pudiera ser notable, y más de un investigador se vería tentado a atribuirlo a su tratamiento; por ejemplo, un grupo de alumnos del segundo año de primaria pueden mostrar gran progreso en su rendimiento en aritmética, después de que se les sometió a un texto programado de aritmética. El investigador podría argüir: "Ya que se observó un cambio mensurable entre la situación anterior a la inclusión de mi método y la situación posterior, entonces se puede inferir que mi método fue el causante del cambio." Este argumento es falaz, aunque no autocontradictorio. Lo falaz del argumento radica en su circularidad: "Si se observó un cambio, éste se debe a mi método." ¿Por qué es efectivo el método? "Porque produjo el cambio observado."

Una manera de superar esta circularidad consiste en buscar un punto de referencia independiente de las dos partes involucradas. Se puede razonar que la VD (las puntuaciones) se puede ver afectada por variables, como el número de horas que cada alumno invirtió frente a su televisión viendo programas como "Plaza Sésamo"; o el desarrollo biológico, o la interacción con hermanos mayores, o el simple paso del tiempo. Todos éstos son puntos de referencia que no dependen del sistema de enseñanza (VI), ni de la medida de ejecución (VD). Tales puntos podrían ayudar al investigador a salir de su circularidad. Si tan sólo el investigador pudiese demostrar que ninguna de estas otras variables afectó a su VD, podría salir de la circularidad de su argumento, demostrando la utilidad de su diseño y la eficacia de su sistema de enseñanza. El problema radica en que él no tuvo ningún tipo de control experimental sobre estas variables. No tuvo, tampoco,

la precaución de dividir a su grupo en dos subgrupos. De haberlo hecho, uno de los subgrupos habría sido sometido al tratamiento, y el otro no. Ambos subgrupos habrían sido medidos antes y después del tratamiento. Sus resultados podrían ser similares a los datos ficticios de la figura 1. En este caso, el aumento en ejecución, logrado en el “grupo

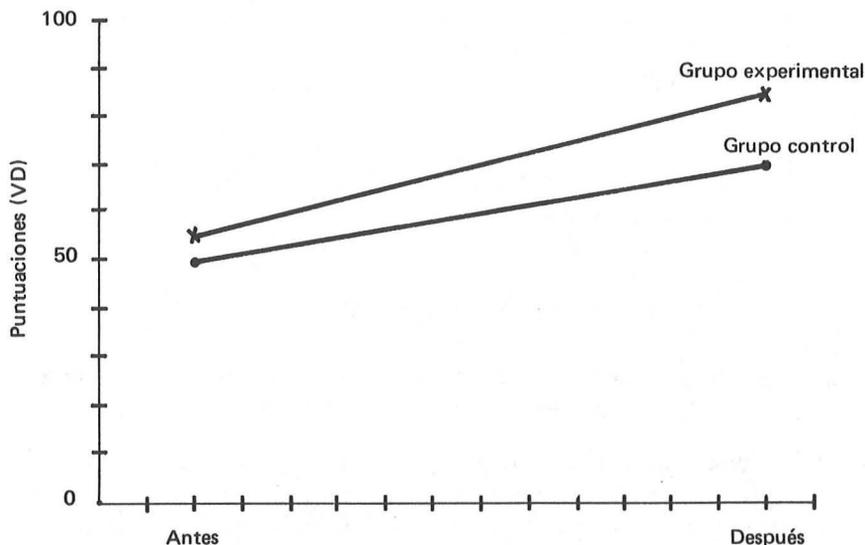


Figura 1. Datos ficticios que muestran un mayor aumento en el grupo experimental que en el grupo control.

experimental”, se podría comparar favorablemente con el aumento observado en el “grupo control”. Estos resultados tenderían a dar algún apoyo a sus afirmaciones acerca de la superioridad de su método. En este punto, parece oportuna la inclusión de una nota precautoria. Aun en el caso de haber incluido un grupo “testigo” (control) como nivel de comparación, la ausencia de técnicas de muestreo (aleatorización) para la asignación de sujetos a los grupos invalidaría al diseño como experimental. La ausencia de control experimental constituiría una fuente de invalidez todavía más importante.

No obstante, sus resultados también pudieran ser similares a los datos ficticios de la figura 2. En este caso, el paralelismo entre la línea que representa al “grupo experimental” y la que representa al “grupo control” indicaría que las variables que están operando comúnmente sobre ambos grupos, serían las principales responsables del cambio observado y no la VI (método). Un tercer caso, el cual sería completamente desalentador para nuestro investigador, lo sintetiza la figura 3.

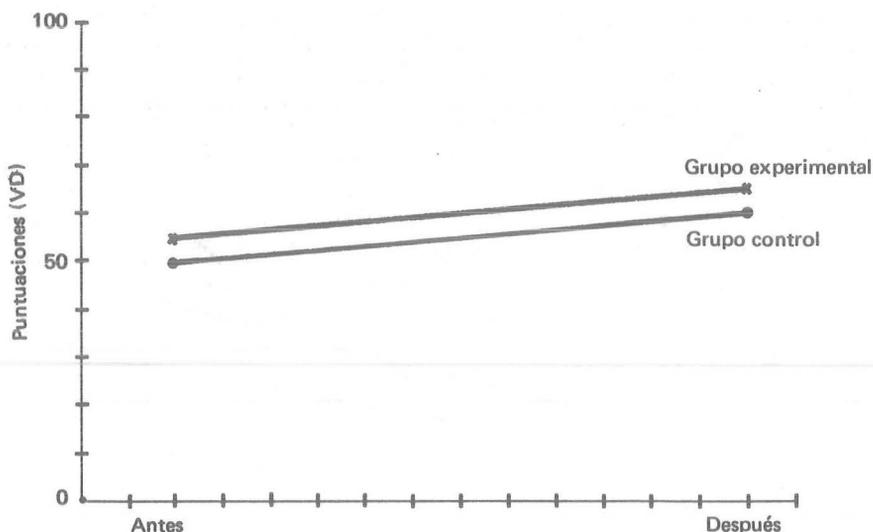


Figura 2. Datos ficticios que muestran un aumento similar entre el grupo control y el grupo experimental.

Aquí, a pesar de que ambos grupos muestran progreso, el “grupo control” estaría cambiando en una forma más drástica en la dirección favorable, que el “grupo experimental”. Recuérdese que nuestro inves-

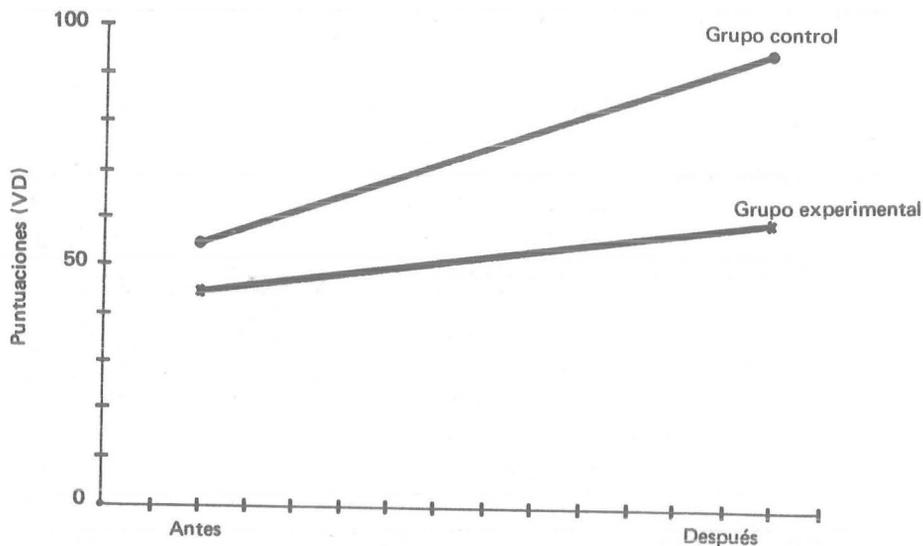


Figura 3. Datos ficticios que muestran un mayor aumento en el grupo control que en el grupo experimental.

tigador imaginario no podría hacer ninguna de estas comparaciones, ya que todo lo que él podría graficar sería la línea "E" de las tres figuras.

Como nuestro hipotético investigador empleó el diseño de un solo grupo mediante pretests y postests, lo único que válidamente puede argüir es que se produjo un cambio (si es que lo hubo) entre la primera y la segunda medición. Fuera de esta limitada afirmación, toda especulación acerca del papel que desempeña la VI en la determinación del cambio sería aventurada. Por una parte, no se podría argumentar positivamente contra la influencia de n variables alternativas como las mencionadas anteriormente; por otra parte, hay el riesgo de caer en una afirmación circular.

VARIABLES EXTRAÑAS

Campbell y Stanley (1963) mencionan un número de variables (VE) que pudieran competir con la VI para determinar un cambio en los valores de la VD en este diseño. Estas variables son: "historia", maduración, medición, instrumentación, regresión estadística, "mortalidad", e interacción entre varias de ellas. Estas variables, por cierto, en su mayoría son las mismas que constituyen serios obstáculos para la validez interna.

La "historia" se refiere a la intervención de eventos que afectan a la mayoría de los sujetos entre la toma de la primera medición y la segunda. El hecho de que se introduzca en la escuela un equipo nuevo de televisión de circuito cerrado, o incluso el cambio de director, son ejemplos de este tipo de evento. Campbell y Stanley señalan que mientras más tiempo transcurra entre las dos mediciones, más probable será que haya algún efecto de esta clase de variable. No debe extrañar, entonces, que una práctica común en psicometría sea la de aplicar la misma prueba (o formas equivalentes) en lapsos relativamente cortos.

La *maduración* se refiere, básicamente, a los procesos de desarrollo biológico que puedan sufrir los sujetos entre ambas mediciones. Si las etapas de "desarrollo intelectual" surgen paralelamente al desarrollo biológico, como Piaget y sus colaboradores parecen sugerir, entonces se podrían producir cambios importantes en la VD debidos a la maduración y no a la VI.

La *medición* se refiere a los efectos reactivos de un determinado "instrumento" de medición. Un instrumento es reactivo cuando produce cambio, al mismo tiempo que mide el cambio. La inmensa mayoría de los tests psicológicos son reactivos. Siempre que se aplica el mismo test

(o su forma equivalente) por segunda vez, se están “recogiendo” los efectos de la primera aplicación. Si los efectos son aditivos, entonces el aumento aparente en los valores de la VD puede atribuirse equivocadamente a la VI. Este problema de reactividad de los instrumentos no es propiedad exclusiva de la psicología y de ciencias afines, sino que sucede también en otras disciplinas científicas. En un número reciente de la revista *National Geographic* (1973) se presenta el minilaboratorio “Vickingo”, destinado a recoger muestras del suelo marciano. El instrumento escarba en el suelo, recoge muestras, las analiza y manda los resultados de los análisis a la Tierra. Es tan sensible que si accidentalmente se cayese una fracción diminuta de pintura, el aparato la analizaría. Los “datos” así obtenidos serían producidos por el instrumento de medición, y no por el evento a ser medido.

La *instrumentación* se refiere a cambios en los instrumentos de medición o de estimulación que pasan inadvertidos al experimentador; por ejemplo, un instrumento de estimulación funcionó correctamente durante la primera situación experimental (encendió cierta luz en el 100% de las veces) y funcionó incorrectamente durante la segunda situación experimental (sólo encendió en un 80% de las veces intentadas). En este caso, habrá una diferencia en las dos mediciones tomadas. Esta diferencia reflejará la falla de la instrumentación; sin embargo, si la falla no fuese advertida por el experimentador, las diferencias obtenidas podrían ser malinterpretadas y se inferiría erróneamente un efecto de la VI sobre la VD. Por otra parte, la falla puede radicar en un instrumento de medición; por ejemplo, en una caja de Skinner para palomas, el disco translúcido que el animal utiliza para emitir su respuesta y obtener un reforzador (un grano de comida) puede tener un resorte flojo y no cerrar un circuito que la conecta a un graficador. Si el investigador ignorase este hecho, podría pensar que la tasa de respuesta del organismo ha disminuido; entonces, podría hacer cualquiera de estas dos cosas: a) apresurarse a atribuir la disminución a la VI, o b) buscar meticulosamente la falla en su equipo programador, en su control sobre el organismo, etc., hasta encontrarla y corregirla.

La *regresión estadística* es un fenómeno estadístico que intrigó a Galton a fines del siglo pasado y que originó toda un área de la estadística moderna. Al respecto, Galton observó que los hijos de padres altos tendían a ser menos altos de lo que se esperaría, a juzgar por la estatura del padre. Los hijos tendían a “regresar” a la media aritmética de la estatura de sus padres (su estatura se hallaba entre la del padre y la de la madre). De su concepto original de “reversión” (Dudycha y Dudycha, 1972), se derivaron los conceptos de

“correlación” y de “regresión estadística”. Este último es, básicamente, un problema relacionado con una baja correlación (poca relación) entre dos variables; en este caso, dos mediciones sucesivas. Mientras más baja sea la correlación, más probable será que ocurran efectos de “regresión”. Estos efectos de “regresión” operan tanto en el extremo alto de la VD como en el bajo. En el último caso, el efecto se puede considerar como de “progresión”, ya que una puntuación baja en la VD en la primera medición se verá aumentada (“progresando” hacia la media) en la segunda medición. Por lo contrario, una puntuación alta se verá disminuida en la segunda medición (“regresará” hacia la media). Estos efectos son, puramente, el resultado de cierto comportamiento estadístico de los datos, y la “falacia de regresión” consiste en darles una interpretación apresurada, en la cual se pueden atribuir los “efectos” al tratamiento que aconteció entre ambas mediciones: “. . . los (estudiantes) brillantes parecen volverse lerdos, y los lerdos más brillantes. . .” (Campbell y Stanley, 1963).

Hays (1963) sintetiza la regresión hacia la media diciendo: “. . . un individuo caerá relativamente más cerca de la media del grupo en la cosa predicha, de lo que lo está en la cosa que realmente se conoce” (Hays, 1963, pág. 501). Señala también que los efectos de regresión hacia la media son resultado de esta forma particular de predicción (lineal), más que un fenómeno del mundo natural.

Además de una baja correlación entre ambas mediciones, hay otros factores, como el error de medición o el “techo” de la escala utilizada, que pudieran producir este tipo de efecto. Este problema, de nuevo, se puede resolver con una aproximación no estadística por medio de un diseño más eficaz y, sobre todo, a través del empleo riguroso del control experimental. No se pueden atribuir efectos a un tratamiento, a menos que se demuestre inequívocamente una relación directa entre la manipulación de la VI y el cambio observado en la VD.

Finalmente, la “mortalidad” se refiere a la pérdida o deserción de sujetos a través de un experimento o investigación. Si un sujeto está participando en un experimento con el fin de obtener puntos en un curso determinado y a medio semestre decide abandonar el curso (y, por tanto, el experimento), entonces contribuirá a crear o aumentar la “mortalidad” del experimento. Asimismo, un animal que, en un intento de escapar de su jaula, se hiere y queda impedido para responder en una situación experimental, también contribuye a la “mortalidad”. Éste es un problema común en un gran número de investigaciones, especialmente en aquellas en que intervienen voluntarios humanos. Si la mortalidad es extrema, entonces se pueden producir datos anormalmente extremos o anormalmente semejantes. En el caso del diseño de un solo

grupo mediante pretest y postest, esta variable pudiera invalidar completamente una investigación; por ejemplo, en el caso de que se incluyera un nuevo sistema de enseñanza, sólo los estudiantes más "motivados" al cambio podrían cooperar durante toda la investigación, y los otros desertar. Probablemente habría un aumento en los valores de la VD, sólo como un efecto de la "mortalidad".

3.2. DISEÑO DE COMPARACIÓN ESTÁTICA

Este diseño preexperimental es uno de los más frecuentemente utilizados en la investigación educativa, social y transcultural. En este diseño, no es el investigador quien determina el modo en que se forman los grupos con los que va a investigar. Cuando los niños de la "escuela 1" reciben cierto tratamiento mientras los niños de la "escuela 2" no lo reciben, entonces se está comparando su ejecución en α test de rendimiento escolar, y se está utilizando este tipo de diseño. Cuando se comparan las "actitudes" de un grupo perteneciente a una clase socioeconómica "alta" con un grupo de una clase socioeconómica "media-baja", también se está haciendo una comparación estática. Finalmente, cuando se comparan los "estilos cognoscitivos" de niños mexicanos con los de niños japoneses y norteamericanos, tenemos otra forma de comparación estática. Lo estático se refiere a que el "grupo", o más correctamente, "la población", ya está formada. A pesar de que el investigador solamente seleccione una muestra de cada población, y no obstante que siga técnicas de aleatorización o estratificación, esto no alterará las variables que determinaron las características de la investigación. Existen dos problemas fundamentales en este diseño: a) cuando se aplica alguna clase de tratamiento, se pueden confundir los efectos de las variables "naturales" con los del tratamiento, y b) cuando no se aplica ningún tratamiento y sólo se comparan dos o más grupos en una o más VD, entonces se pueden atribuir las diferencias encontradas a una "variable-ómnibus" o "supervariable", como son la cultura, la clase social o la personalidad.

Una fuente de invalidez interna particularmente relevante a este diseño es la *selección*. Ésta se refiere a la aplicación de un procedimiento no experimental para la asignación de sujetos a los grupos o condiciones experimentales. Cuando se emplea el diseño de comparación estática con estudiantes de dos diferentes universidades, la selección puede ser una variable fundamental para la invalidación del estudio. Se sabe que diferentes universidades utilizan distintos procedimientos de admisión. Algunas requerirán que sus estudiantes se sometan a un mayor número de exámenes de admisión que otras;

algunas tendrán un criterio flexible que permita la entrada de estudiantes con un bajo promedio de educación preuniversitaria, mientras que otras sólo admitirán a aquellos estudiantes que tengan promedios más altos. Todas éstas son formas de selección.

La selección está determinada por una serie de criterios no experimentales, y el investigador no tiene nada que hacer en dicho proceso. Cualquier comparación que se haga entre estudiantes de dos o más universidades con diferentes criterios de admisión, implica un proceso de selección que podría ser determinante de las diferencias observadas en la VD. El riesgo de que se presenten problemas de selección se extiende a todos aquellos casos en que un investigador aproveche la circunstancia de que ciertos grupos o "clases" se hayan formado "naturalmente", sin su intervención.

3.3. ALGUNOS DISEÑOS CUASIEXPERIMENTALES

Aun cuando Campbell y Stanley (1963) consideran a los diseños cuasiexperimentales como un punto intermedio entre los diseños preexperimentales y los experimentales, en este libro se estudiarán como parte de los diseños preexperimentales. De acuerdo a la regla de clasificación expresada en el principio de este capítulo, aquellos diseños en los cuales el investigador no pueda ejercer un completo control experimental sobre las variables implicadas (manipulación de las VI, eliminación o minimización de las VE y adecuado registro de las VD), serán considerados como preexperimentales.

Un diseño cuasiexperimental puede conceptualizarse como una clase particular de diseño preexperimental. Sus principales características son: *a*) el empleo de escenarios naturales, generalmente de tipo social; *b*) la carencia de un control experimental completo; específicamente, la imposibilidad para controlar una o varias clases de VE; *c*) el uso de procedimientos —como el de aleatorización o el de producir múltiples observaciones— como sustitutos del control experimental. La finalidad de dichos procedimientos consiste en minimizar (y en algunos casos, eliminar) los efectos de tantas fuentes de invalidez interna como sea posible, y *d*) su disponibilidad; es decir, por una parte, pueden utilizarse cuando no sea posible emplear un diseño experimental, y por otra, algunas veces pueden explotar la conformación de alguna situación social. Algunos de estos diseños se presentan a continuación. [Para un estudio más extenso, se remite al lector al trabajo de Campbell y Stanley (1966).]

3.3.1. Diseño de series de tiempo

En un sentido general, una serie de tiempo se refiere a la obtención de un conjunto de mediciones u observaciones tomadas en diferentes periodos. El estudio de series de tiempo es fundamental en disciplinas como la economía y la sociología (véase, por ejemplo, Yamane, 1970). En la gran mayoría de los estudios económicos o sociológicos en que se utiliza el análisis de series de tiempo, el investigador no determina los cambios que ocurren en la serie; por ejemplo, las observaciones pueden ser acerca del producto nacional bruto de determinado país, o referirse a índices inflacionario o de mortalidad; en ninguno de estos casos podría un investigador determinar que los índices aumentarían o disminuirían. El investigador carece del poder para alterar la(s) VI, de la(s) cual(es) las diferentes VD son una función.

El diseño de series de tiempo, en la investigación psicológica, tiene la característica de que el investigador puede introducir una VI con el propósito de alterar la forma de la serie. La inclusión de la VI generalmente se ve precedida de un número de observaciones y seguida de otra cantidad de observaciones. La forma de la línea que representa a la serie de observaciones a través del tiempo, puede depender básicamente de variables ajenas a la VI; por ejemplo, en la investigación económica se reportan cambios o fluctuaciones estacionales (de acuerdo a las estaciones del año) o cíclicas. Ambos tipos de fluctuaciones se presentan con regularidad y pueden depender de la forma del año fiscal de un gobierno, de los periodos arbitrarios de vacaciones y días festivos de un país o región, de la forma en que las universidades dividen sus semestres, etc. Cualesquiera que sean las causas del cambio, éstas se hallan fuera del control del experimentador.

Cuando se utiliza el diseño de series de tiempo, se trata de evitar, hasta donde sea posible, la influencia de fluctuaciones estacionales o cíclicas. El hecho de haber realizado algunas mediciones u observaciones anteriores a la inclusión de la VI, puede servir para evaluar cuál es la tendencia de los datos antes del tratamiento. Cualquier efecto de la VI deberá apartarse considerablemente de dicha tendencia. Los efectos de la VI podrían ser inmediatos y a corto plazo, en cuyo caso la tendencia de los datos variaría repentinamente en el punto donde se aplicara la VI, y la tendencia general sería paralela a la tendencia anterior. Las observaciones posteriores al tratamiento resultan útiles, entre otras cosas, para hacer esta comparación. El efecto de la VI puede ser retardado o a largo plazo. En el primer caso, podrían obtenerse varias observaciones posteriores a la inclusión de la VI no desviadas de la

tendencia previa, y, pasadas estas observaciones, aparecería un notorio cambio en la tendencia. En el segundo caso, la tendencia previa al tratamiento debería cambiar al presentarse éste (la VI), y permanecer en esta nueva "dirección" durante el resto del tratamiento. En todos los casos, mientras mayor sea el número de observaciones (tanto anteriores como posteriores), más fácilmente se podrá evaluar el efecto de la VI.

El hecho de realizar múltiples observaciones no sustituye adecuadamente al control experimental, pero minimiza el efecto de algunas de las variables competitivas o fuentes de invalidez interna. Campbell y Stanley (1966) arguyen que este diseño puede controlar la mayoría de las fuentes de invalidez interna que ellos mismos mencionan. Aun cuando puede estarse de acuerdo con ellos respecto a la eficacia del diseño para evitar los efectos de las variables de medición, instrumentación y regresión, el resto de factores de invalidez interna aún puede desempeñar un papel importante. Por ejemplo, los efectos de la maduración pueden observarse en una forma notoria en sólo unas cuantas semanas (como en niños muy pequeños y en la transición pubertad-adolescencia). Los efectos de selección y mortalidad son, más bien, independientes del diseño, y es más probable que ocurran en una situación de investigación social o educativa a largo plazo, que en una situación experimental de unas cuantas sesiones.

Campbell y Stanley sugieren el empleo del diseño en aquellas situaciones en que una institución (una escuela o un hospital) lleva a cabo registros de los sujetos institucionalizados, como parte de su procedimiento regular.

El diseño de series de tiempo para obtener una representación típica del comportamiento de los sujetos antes y después de un tratamiento, es obviamente superior al diseño de un solo grupo con pretests y postests. Aunque el diseño carece de un punto de comparación independiente del sujeto o grupo de sujetos observados y de un control experimental completo, puede proporcionar información exploratoria sumamente útil para el planeamiento de un verdadero experimento o para una mejor comprensión del comportamiento de una VD a través del tiempo. Mediante este diseño, como con los otros diseños cuasiexperimentales, es prácticamente imposible descartar explicaciones alternativas (efectos de VE no controladas) en la evaluación del efecto de la VI sobre la VD.

3.3.2. Diseño de muestras de tiempo equivalentes

Campbell y Stanley (1966) consideran a este diseño como una variante del diseño de series de tiempo. En este tipo de diseño también se mide u observa repetidamente al mismo sujeto o grupo de sujetos. Las diferencias básicas con el diseño anterior son: *a*) cada observación o medición se realiza inmediatamente después de la presentación (o no presentación) de la VI; es decir, no hay un número relativamente grande de observaciones antes de la inclusión de la VI, ni después; *b*) la presentación de la VI o su no presentación ocurren aleatoriamente, de acuerdo a un procedimiento de muestreo, y *c*) es más apropiado emplear este diseño en aquellas situaciones en que se sabe (o se supone) que los efectos de la VI son reversibles y transitorios. Webb y sus colaboradores (1966) revisaron varios estudios en situaciones industriales y en guarderías, donde se utilizaron procedimientos de muestreo de tiempo. En el primer caso, el énfasis del procedimiento radicó en determinar con cuáles variables se relacionaba la baja de actividad a ciertas horas del día. En el segundo caso, era indicado por la gran periodicidad en la conducta de niños pequeños. Obviamente, en ninguno de los dos casos se manipuló ninguna VI. De acuerdo a la clasificación de Campbell y Stanley, un diseño en el cual se empleara el muestreo de tiempo pero en el que no incluyera una VI, no podría considerarse como cuasiexperimental, sino como preexperimental. En el presente libro, este tema no está a discusión ya que tanto los diseños considerados por dichos investigadores como cuasiexperimentales, como los que se han considerado preexperimentales, caen bajo la categoría de preexperimentales. Campbell y Stanley demuestran que este diseño es superior al de series de tiempo en lo que respecta al control de la variable "historia". Señalan que las mayores desventajas del diseño son: *a*) la reactividad de las situaciones en las cuales se utiliza (la presencia de los investigadores produce cambios en la VD por sí misma), y *b*) lo que ellos denominan "interferencia de los tratamientos múltiples". Esta interferencia se refiere a los efectos retardados que un tratamiento pudiera tener y que se podrían reflejar en las medidas de la VD, obtenidas durante periodos de no tratamiento. Además, indican que el análisis de los datos podría verse invalidado o alterado considerablemente por los efectos del error de muestreo de ocasiones. Por último, sugieren que cuando se utilice el mismo tipo de VI (estímulo) sobre el mismo sujeto, se deberá evitar todo tipo de periodicidad en la estimulación. Esto sólo sería posible si el número de mediciones fuese lo suficientemente grande.

3.3.3. Diseño de comparación estática con pretests

Este diseño puede ser considerado como una extensión del diseño de comparación estática. En el libro de Campbell y Stanley (1963), este tipo de diseño se denomina "diseño de grupo control no equivalente".

Las características principales de este diseño son: *a*) no hay equivalencia entre los grupos antes del tratamiento, ya que no se utiliza ningún proceso de aleatorización para asignar los sujetos a los grupos; en vez de eso, se toman dos grupos ya formados de acuerdo a procedimientos no experimentales; y *b*) se incluye un pretest; es decir, se hace una observación previa al tratamiento en ambos grupos. Esto no basta para hacer equivalentes a los grupos ni para sustituir al control experimental, pero al menos da un indicio de qué tan diferentes son los grupos al empezar.

Pueden hacerse varias comparaciones con este diseño. Las diferencias obtenidas en dichas comparaciones pueden producir interpretaciones erróneas acerca del efecto de la VI. La primera comparación se efectúa entre los valores de la VD antes del tratamiento (pretest); aquí, el investigador tratará de obtener la menor diferencia posible entre sus dos grupos. Mientras menor sea esta diferencia, mayor será su confianza en que se está aproximando a una situación de control experimental. Si la diferencia es grande, entonces le será más difícil evaluar sus resultados.

La segunda comparación se lleva a cabo entre las puntuaciones de la VD antes y después del tratamiento en el grupo experimental. Las mismas precauciones que se señalaron en relación al diseño de un solo grupo con pretests y postests se observan en este caso. Si un investigador estuviese comparando los efectos de un texto programado con uno de tipo común, no debería clamar la efectividad del texto programado solamente en base a esta comparación. Los efectos observados (un aumento en las puntuaciones de la VD) se pueden deber a diversas variables, además de la VI (texto programado). Entre éstas, tenemos la historia, la maduración, la selección y la medición.

También podría suceder que el grupo control presentara un aumento en la VD, en cuyo caso habría la tercera comparación posible. Si ambos grupos aumentan, esto hace más dudoso el efecto de la VI. El aumento en el grupo experimental debería ser sumamente notable, en comparación con el incremento en el grupo control para poder atribuir el efecto del aumento a la VI. Campbell y Stanley señalan que, en ciertas ocasiones, una interacción entre algunas de las variables selección, madu-

ración, medición e historia podría producir una interpretación inadecuada y, por tanto, una fuente de invalidez interna. Un ejemplo claro de este tipo de interacción es el siguiente: supóngase que se compara un grupo de los mal llamados “enfermos mentales” con un grupo control (visitantes de un hospital psiquiátrico). Supóngase también que se “mide” tanto a “enfermos” como a visitantes mediante un test de “salud mental”. Los visitantes (seguimos suponiendo) obtienen puntuaciones cerca del límite de la escala de “salud mental”. Por otra parte, los “enfermos” se encuentran ante una situación de prueba que les ha sido molesta en el pasado, y además desconocen este test. Obtienen puntuaciones bajas en el pretest. Para la segunda ocasión, después de algún tratamiento, los pacientes ya no reaccionan en la misma forma a la situación y, además, ya están familiarizados con el test; esta vez obtienen puntuaciones altas, cerca de los límites de la escala. Por otra parte, los visitantes obtienen puntuaciones similares a las de la primera ocasión (no progresan).

Ahora bien, un investigador poco cuidadoso atribuiría el “cambio observado” al tratamiento que intervino entre las mediciones, ignorando así una evidente interacción entre la selección y la medición. Esta clase de problemas no puede evitarse mediante este diseño, aunque es menester estar alerta para cuando se presenten.

La cuarta comparación se puede efectuar entre las puntuaciones de ambos grupos en la VD después del tratamiento. Una diferencia (por ejemplo, a favor del grupo experimental) en este punto no es interpretable por sí sola. Si, por ejemplo, hubo una diferencia en el pretest y ahora se obtiene una diferencia semejante, sería difícil argüir algo a favor de la VI; sin embargo, si no hubo diferencias en el pretest, hubo un incremento notable en el grupo experimental comparado con el aumento del grupo control, y además se observó una diferencia importante en el posttest a favor del grupo experimental; entonces, en este caso hay bases más firmes para pensar en un verdadero efecto de la VI. Una representación típica de esta situación se muestra en la figura 4.

Por último, la quinta comparación se puede realizar entre las diferencias del pretest y las del posttest. De las cinco comparaciones consideradas individualmente, ésta es probablemente la más segura de todas. En primer lugar, esta comparación da cuenta de una posible disparidad (no equivalencia) de los grupos antes del tratamiento. Aun siendo éste el caso, el tamaño relativo de la segunda diferencia (el posttest comparando los dos grupos), comparada con la primera, es un indicador bastante preciso de los efectos producidos por la VI. De cualquier manera, un panorama general en el que se incluyan estas cinco compa-

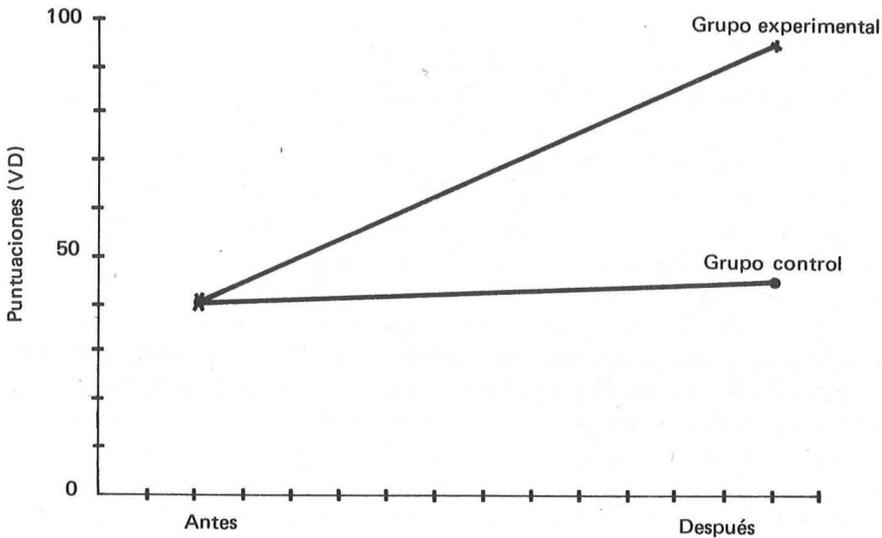


Figura 4. Datos ficticios que muestran un claro efecto de la VI sobre la VD.

raciones será la mejor forma de evaluar los efectos de la VI usando este diseño preexperimental.

3.4. DISEÑOS APAREADOS-CORRELACIONALES

En la literatura experimental es común encontrar el diseño de grupos indirectamente apareados (apareados a través de una r), como parte de los diseños experimentales. En esta sección no se le considerará como un diseño experimental, ya que no presenta la característica de demostrar que los grupos "apareados" son equivalentes antes de la aplicación de la VI.

Algunos diseños experimentales deben producir situaciones en las cuales las observaciones previas a la inclusión de la VI sean equivalentes entre sí. En el caso de que el mismo sujeto o grupo de sujetos sean sometidos a una comparación, se debe obtener un número de observaciones lo suficientemente grande y estable como para evitar confusiones debidas a información incompleta. Cuando se utilizan dos o más grupos de sujetos, entonces debe haber algunas indicaciones de que los grupos son equivalentes. Se deberían controlar los efectos de la selección, así como los de la subjetividad del experimentador, al asignar sujetos a determinadas condiciones. También se debería observar o medir a los sujetos de los diferentes grupos antes de la aplicación

de la VI, asegurándose de que un grupo no era superior a los otros en la VD desde un principio. Un procedimiento común en la mayoría de los diseños experimentales tradicionales (de grupos) es el de la aleatorización para obtener la equivalencia entre los grupos.

Un procedimiento experimental que garantiza la equivalencia entre los grupos es el del *apareamiento directo*, es decir, aquel procedimiento en que se aparean los sujetos de las diversas condiciones experimentales de acuerdo a los valores de la VD en una medición original. Nótese que en este procedimiento se utiliza la misma VD que se trata de afectar con la VI. Este es un apareamiento directo y una forma que califica a un diseño como perfectamente experimental. Por otra parte, tenemos los procedimientos de *apareamiento indirecto* o *correlacional*, los cuales producen equivalencias poco convincentes entre los grupos apareados.

En el diseño de grupos indirectamente apareados, se parte de la convicción o fuerte suposición de que si se mide a los sujetos en la VD antes del tratamiento, se producen efectos reactivos de la medición, que podrían estar presentes aún en la segunda medición. Para evitar estos efectos, el investigador debe usar este diseño basándose en una de las siguientes alternativas: *a*) una suposición, o *b*) una correlación. Cuando se basa en una suposición, simplemente asume que si mide a los sujetos en una tarea altamente relacionada (variable de apareamiento) con la VD y los aparea de acuerdo a esta segunda variable, entonces está obteniendo equivalencia entre los grupos. Para llegar a esta suposición, el investigador puede partir en el peor de los casos, desde el simple "sentido común", o en el mejor de los casos llevar a cabo un estudio exploratorio para determinar el grado de correlación entre la VD y la VA (variable de apareamiento). Esto nos lleva a considerar la segunda alternativa. Cuando la suposición se deriva de una o más correlaciones entre la VD y la VA, el investigador debe revisar la literatura referente a su estudio y averiguar qué tanto se relacionaron ambas variables en el pasado. La cuestión no es tan simple cuando advierte, para su sorpresa, que las correlaciones encontradas entre ambas son bajas. En este caso, le queda la alternativa de buscar una segunda, tercera o cuarta VA, hasta encontrar aquella que, de acuerdo a la literatura revisada, tenga una alta correlación con la VD. Aun en el caso de que encontrara efectivamente una VA que tuviera una alta correlación con la VD, el experimentador se enfrentará al problema de que las correlaciones "se mueven"; es decir, una correlación no es un índice estable del comportamiento de dos variables. Así, dos variables que anteriormente tuvieron una alta correlación, después pueden tener una correlación de cerca de cero, y más tarde pueden tener una correlación moderadamente alta. Toda correlación, a menos que

su valor sea prácticamente de 1.0, no garantiza que una variable sea equivalente a otra; incluso, cuando "dos variables" presentan una correlación de 1.0, indican que se están asignando dos nombres diferentes a una misma cuestión.

También puede presentarse el caso de que, después de buscar afanosamente y no encontrar la VA en la literatura, el investigador decida hacer estudios exploratorios para determinarla. Estos estudios podrían o no tener éxito, pero es seguro que el esfuerzo que implican no se justifica, en vista de las alternativas a disposición del investigador. ¿Cuáles son esas alternativas? Por una parte, está la posibilidad de utilizar un procedimiento de medida no reactivo, lo que permitiría un apareamiento directo; por otra, está el conocido recurso de la aleatorización.

4

conceptos básicos del diseño experimental

UNA DE LAS concepciones más completas acerca de lo que es un diseño experimental pertenece a Kirk (1972). Este autor considera que, entre otras cosas, un diseño experimental es un plan de acuerdo con el cual se asignan los sujetos a los diferentes grupos o condiciones experimentales. Esta concepción se aplica perfectamente a los diseños experimentales tradicionales o de grupo. En ellos, el plan (el diseño) determina de antemano cuántos grupos de sujetos se deberán incluir, cuántos sujetos deberá haber en cada grupo y qué tipo de análisis estadístico es el más indicado. Sin embargo, habrá ocasiones en que únicamente se trabaje con un solo sujeto. Entonces, en este caso, no tendría sentido hablar de "asignarlo" a un determinado grupo. Al ampliar Kirk el concepto de grupo al de condición experimental, ayuda a la inclusión de los diseños para organismos individuales a su concepción.

El concepto de condición experimental se refiere a un valor o combinación de valores de una o más VI. Así, por ejemplo, si un experimento requiere la formación de un grupo experimental y un grupo control (testigo), entonces se puede decir que el diseño experimental consiste de una condición experimental y de una condición control. Asimismo, si el experimento requiere que un mismo sujeto pase por una etapa de

observación y luego por una de tratamiento, entonces se puede decir que el diseño consiste de una condición de observación y de una condición de tratamiento. Es obvio que el experimentador que trabaja con grupos de sujetos encontrará más fácil referirse a grupos experimentales o controles. El experimentador que trabaja con organismos individuales, por su parte, hallará más fácil hablar de línea base y manipulación de la VI. Pero si se ha de utilizar un lenguaje más general para referirse a los diseños experimentales, el término genérico de "condiciones" parece llenar el requisito de describir adecuadamente a ambas aproximaciones. Ahora bien, si un solo sujeto va a ser sometido a varias condiciones experimentales, entonces la operación de asignarlo a ellas pierde un tanto su significado. Lo que se vuelve importante como característico del diseño no es la operación de asignación de sujetos a condiciones experimentales, sino la estructura misma del plan, a saber: el arreglo de condiciones. Las condiciones se disponen de acuerdo a la pregunta de investigación, a los valores de la VI que se van a manipular y al tipo de metodología seleccionado. Algunas veces, la aplicación de un arreglo de condiciones (diseño) se ve restringida por condiciones ajenas al investigador, en cuyo caso se le debe sustituir por un arreglo diferente o más simple. Este es un problema de selección del diseño que será estudiado más adelante. Por el momento, analizaremos algunos puntos básicos de dos diferentes tipos de metodología que son determinantes para la creación, selección o uso de distintos diseños experimentales. Dichos puntos básicos son: *a*) la diferencia entre experimentos con grupos de sujetos y experimentos con organismos individuales; *b*) el concepto de grupo control; *c*) el concepto de línea base; *d*) la suposición de equivalencia entre un sujeto medido en numerosas ocasiones y múltiples sujetos medidos una sola vez; *e*) $N = 1$, y *f*) la comparación entre línea base y grupo control.

4.1. EXPERIMENTOS CON GRUPOS DE SUJETOS Y EXPERIMENTOS CON ORGANISMOS INDIVIDUALES

En diversos campos de la psicología (aprendizaje, motivación, formación de conceptos, percepción, memoria, etc.), es posible encontrar un número sustancial de informes relativamente recientes (incluso en la década de 1960 y principios de la de 1970), en los cuales se llevaron a cabo experimentos con grupos de sujetos. Esta ha sido la aproximación tradicional en la psicología experimental (paradójicamente, en ésta primeramente se efectuaron experimentos con organismos individuales), que ha dominado este campo durante lo que va del presente siglo.

Quizá las dos excepciones más importantes a esta regla sean las áreas de la psicofísica y del condicionamiento operante, en las cuales la metodología utilizada indica (prácticamente exige) la ejecución de experimentos con organismos individuales y la toma de múltiples medidas sobre el mismo sujeto. Fuera de la psicología, en el área relacionada con la investigación fisiológica, también es fundamental la metodología individual.

Algunas razones por las cuales se realizan experimentos con grupos de sujetos son las siguientes: *a*) por seguir la indicación derivada de la teoría de la medida en el sentido de que mientras mayor sea el número de observaciones, más se acercarán los datos obtenidos a los "valores verdaderos" del fenómeno medido (por ejemplo, un sujeto podría recordar un número de símbolos excesivamente grande, y sería riesgoso considerar que todos los demás lo pudieran hacer); *b*) por la influencia de la estadística en la experimentación, la cual requiere que se cumplan ciertas suposiciones (normalidad de las distribuciones, homogeneidad de las varianzas, etc.), las cuales a su vez tienen una mayor probabilidad de cumplirse cuando hay un gran número de sujetos en cada condición; *c*) por el tiempo limitado de acceso a los sujetos —por ejemplo, cuando los sujetos son estudiantes de determinada escuela y no es posible interrumpir sus actividades durante largo tiempo—; *d*) porque el tipo de diseño lo requiera —esto es cierto para todos los diseños tradicionales o de grupo—; *e*) por un intento de que los datos sean representativos de la población estudiada; *f*) para producir una replicación intergrupos, y *g*) cuando se desarrolla investigación aplicada cuyo objetivo es resolver problemas de grupos de sujetos —por ejemplo, para desarrollar cursos de instrucción tutorizada como el de Fred S. Keller (véase Ulrich y colaboradores, 1972; Stice, 1971).

Antes de referirnos a los experimentos con sujetos individuales, es importante aclarar que en un gran número de experimentos con grupos de sujetos, cada sujeto es tratado individualmente. Una sesión típica podría consistir de dos experimentadores, atendiendo cada uno a un sujeto diferente en cubículos separados. Después se podría repetir el procedimiento con otros dos sujetos, y así sucesivamente. Estos experimentos se consideran de grupos, porque es el grupo y no el sujeto particular el que se concibe como *unidad experimental* (aquella a la cual se aplica un determinado tratamiento). Los datos de los sujetos individuales se agrupan de acuerdo a la condición experimental a que pertenecen, y son estudiados como pertenecientes a la misma unidad.

Algunas razones por las cuales se efectúan experimentos con organismos individuales son las siguientes: *a*) Por necesidades de control

experimental. Cuando el control de la conducta del sujeto y del ambiente que lo rodea exige cierto comportamiento de parte del experimentador (tareas de registro, control de variables extrañas, programación de la situación experimental, interacción con el organismo estudiado, etc.), tan demandante que resultaría muy difícil trabajar con muchos sujetos. Cuando todas estas dificultades se resuelven mediante el empleo de equipo automático, aún queda el problema del control del organismo estudiado; de este modo, es más fácil controlar a un sujeto aislándolo de otros sujetos, que controlar varios sujetos en contacto unos con otros. La mera interacción entre varios sujetos es una VE que podría descalificar a un experimento. Incluso, la mera presencia de otro organismo de la misma especie o del experimentador puede ser una variable importante que afecte la conducta del organismo estudiado; *b*) cuando el experimentador decide hacer una replicación directa intrasujeto; *c*) por limitaciones de equipo (no se puede trabajar con más de un sujeto a la vez; *d*) por limitaciones de espacio, por ejemplo, cuando no es posible alojar más que a un número limitado de animales o cuando la cantidad de cubículos para experimentar con humanos es restringida, y *e*) cuando se está estudiando un sujeto con características peculiares (véase la sección 4.5).

Esta diferencia entre la experimentación individual y la de grupo es importante, ya que tipifica claramente las dos metodologías contemporáneas en psicología: *a*) la metodología tradicional o de grupo, y *b*) la metodología del AEC (o de la línea base). Por tanto, esta división será mantenida para el posterior estudio y clasificación de diversos diseños experimentales a través de este libro.

4.2. EL CONCEPTO DE GRUPO CONTROL

Este concepto es uno de los más interesantes entre los estudiantes de nivel introductorio en el campo de la psicología experimental. Parece como si el estudiante nunca olvidara que en algunos experimentos hay un grupo control. En esta sección no nos conformaremos con tan limitada información y trataremos el concepto con alguna extensión.

En algunas ocasiones, un grupo control también se conoce como *grupo testigo*, con lo cual se intenta describir su función como "no participante". Para algunos autores, como Kirk (1968), el grupo control simplemente representa un nivel de tratamiento al cual son asignados algunos sujetos. Más específicamente, este nivel se refiere a la condición experimental en la cual hay una ausencia de tratamiento o VI. Por ser el concepto de grupo control o testigo intuitivamente sim-

ple, parece ser común que se le defina (o sólo se le mencione) y se dé por entendido. Veamos con detenimiento algunas de las definiciones.

A veces se encuentra que la definición de grupo control es: "aquel grupo en el cual se aplica un valor o cantidad cero de la VI". Esta definición presenta el siguiente problema: hay variables que para alcanzar un valor de cero, tienen que ser manipuladas afectando a los sujetos (grupo control), a los cuales se pretende no afectar; por ejemplo, si la VI fuese temperatura en grados C y se pretendiera estudiar el efecto que ejerce el aumento de la temperatura en la ejecución de un tipo de paloma, entonces al aplicar un valor cero de la VI (temperatura) al grupo control, se estaría implicando la manipulación de la VI, afectando considerablemente de esta manera al supuesto "grupo control". En este caso, realmente no se tendría un grupo control, sino dos grupos experimentales: al que se le aumentó la temperatura y al que se le disminuyó a cero.

Hay algunos casos en que sí tiene sentido hablar de un nivel cero de la VI en relación al grupo control. Si, por ejemplo, la VI fuese "número de horas de privación de comida", los grupos experimentales podrían tener 16 y 24 horas de privación, mientras que el grupo control tendría cero horas de privación. Pero si la VI fuese "cantidad de calorías consumidas" por el organismo antes del experimento, entonces serían los grupos experimentales los más cercanos al valor cero de la VI, y no el grupo control, lo cual haría a la definición insostenible. La definición de Kirk evita este problema al referirse, en general, a "ausencia de tratamiento" y no específicamente a un "valor cero de la VI". Sin embargo, el mencionar la ausencia de tratamiento sin referirse al resto del procedimiento también deja sus dudas respecto a la definición de Kirk.

Johnson y Solso (1971) definen al grupo control como "aquel... que es tratado exactamente igual que el grupo experimental, excepto que no recibe el tratamiento experimental". Esta definición, evita los problemas de las anteriores, puesto que incluye una mención de cómo son tratados los sujetos y la ausencia de la VI. Un grupo control, entonces, tiene la función de servir como punto o nivel de comparación para evaluar los efectos de la VI sobre el grupo experimental. Ambos grupos son sometidos al mismo tipo de control experimental, a fin de evitar el efecto de las VE; a los dos se les aplica el mismo procedimiento, y ambos son medidos por el mismo equipo en la(s) misma(s) VD; de este modo, si el tratamiento (VI) ejerce algún efecto sobre la VD, el grupo experimental deberá producir resultados marcadamente diferentes a los del grupo control.

Sería conveniente analizar aquí el hecho de que se están efectuando múltiples investigaciones en un gran número de laboratorios, donde se utilizan varios grupos control. Por ejemplo, hay trabajos acerca de los efectos que producen inyecciones intracerebrales de ARN (véase Gorowitz, 1969) en la memoria, en algunos de los cuales se inyectaron soluciones salinas (condición control del efecto de inyectar en el cerebro, en general) a un grupo control de animales, mientras que a otro grupo control no se le inyectó nada. Gorowitz también informa del trabajo de Corning y John en el condicionamiento clásico de planarias, en el cual se utilizaron dos grupos experimentales (dos formas de administración de ARN) y ocho grupos control, a fin de evaluar los efectos de la sensitivización, el tiempo transcurrido entre el condicionamiento y la medición, el tiempo de regeneración, etc. El mismo Gorowitz revisó un estudio similar de Hartry y sus colaboradores, en el cual se utilizaron cinco grupos control. Los tres primeros estaban ligados a los grupos experimentales y aunque no fueron condicionados, sí fueron manipulados en una forma semejante a las planarias experimentales, y uno de ellos sólo recibió la luz que supuestamente actuaba como estímulo condicionado. Los otros dos grupos control sirvieron de la siguiente manera: uno como recipiente de planarias no tratadas (para observar el efecto separado de la ingestión de unas planarias por otras), y el otro como no recipiente. Este estudio representa cierta importancia en su campo porque demuestra que cualquier tratamiento dado a las planarias donantes (ya sea la simple manipulación, la luz o el choque), además del condicionamiento, afectó a las planarias recipientes haciendo que respondieran mejor.

El empleo de múltiples grupos control implica una mayor restricción de la definición postulada por Johnson y Solso. La definición restringida es: "un grupo control es aquel que es tratado exactamente igual que el (o los) grupo(s) experimental(es), excepto que no recibe alguna de las operaciones que componen la aplicación de la VI o tratamiento", y será la que emplearemos en este libro.

4.3. EL CONCEPTO DE LÍNEA BASE CONDUCTUAL. UN SUJETO COMO SU PROPIO CONTROL

En términos generales, una línea base es un registro estable y sensible de la frecuencia con que ocurre una determinada clase de respuestas durante un periodo arbitrario.

Hay varios conceptos íntimamente relacionados con el de línea base que es conveniente discutir, antes de proseguir con esta exposición.

Skinner y sus colaboradores (Skinner, 1938, 1953, 1969; Ferster y Skinner, 1957; Holland y Skinner, 1970), repetidamente y bajo diferentes enfoques, han enfatizado que el dato básico de la conducta es la *tasa de respuesta* (o, en general, la probabilidad de ocurrencia de una respuesta).

La tasa de respuesta es, simplemente, el número de respuestas en una determinada unidad de tiempo. Ferster (1953) arguye que prácticamente todos los problemas de la ciencia de la conducta se relacionan con este dato básico. Las preguntas acerca de la probabilidad de respuesta, su magnitud y su separación temporal, constituyen ejemplos de dichos problemas relacionados. Es importante señalar que la ocurrencia de una clase de respuestas, tal como se le maneja en la metodología de la línea base, es relativamente "libre"; es decir, el experimentador no impone restricciones acerca de cuándo "debe" ocurrir una respuesta o cuál es el tiempo límite para alcanzar un criterio arbitrario. La conducta que ocurre bajo estas circunstancias es conocida como *operante libre*.

El mismo Ferster considera a la operante libre como un método de gran generalidad que "...se refiere a cualquier aparato que genera una respuesta cuya ocurrencia requiere poco tiempo y deja al animal en el mismo sitio, listo a responder de nuevo" (1953, pág. 263).

La línea base es un término genérico que incluye tanto al de tasa de respuesta como al de operante libre. Una línea base se puede obtener mediante un registro continuo del tiempo, como sucede en un registro acumulativo, o por medio de un registro discreto del tiempo, como ocurre en estudios de modificación de conducta (ACA). Una línea base puede tomarse antes, durante y después de la aplicación de una VI.

La línea base antes y después del tratamiento se puede considerar como línea base observacional, ya que el experimentador no interviene en absoluto para determinar la forma que ésta adquiere. La línea base que ocurre como producto de un tratamiento es considerada como manipulativa, ya que en este caso el experimentador manipula ciertas variables, a fin de producir cambios notables en la línea base. Es posible (y útil) obtener líneas base durante la transición de un estado conductual a otro o durante un proceso de adquisición. Recientemente, Thompson y Grabowski (1972) diferenciaron entre línea base y nivel operante, dependiendo de si la respuesta ya se encuentra en el repertorio del organismo o si es condicionada experimentalmente. En este volumen utilizará el término *línea base* para referirse a ambos.

Sidman (1960) señaló que las principales características de la línea base son las siguientes:

v a) *Estabilidad*. Una línea base, para ser útil, debe ser estable. La conducta del sujeto debe alcanzar un estado fijo, a fin de poderse utilizar como nivel de comparación. Conseguir un estado de estabilidad de determinada respuesta es, básicamente, un problema de control. Es necesario identificar las variables de las cuales una respuesta es una función y, entonces, proceder a establecer un control sobre ellas, con el fin de producir precisamente el tipo de comportamiento deseado. Cuando el investigador no tiene absoluto control sobre las variables que determinan una respuesta (como sucede en múltiples situaciones del ACA), entonces deberá registrar la línea base durante un periodo controlado hasta que esté absolutamente seguro de que la conducta se ha estabilizado, antes de proceder a aplicar un tratamiento.

En numerosas ocasiones, la conducta bajo estudio producirá una línea base recuperable o reversible. Esto significa que después de que se aplica el tratamiento, al darse éste por terminado, la línea base vuelve a su nivel original (previo al tratamiento). Cuando esto sucede, se dice que estamos en presencia de un proceso reversible. Sidman señala que cuando se logra una línea base estable y un buen control de las variables que determinan una respuesta, entonces se puede hacer una completa demostración, en la que el experimentador "activa" y "desactiva" la conducta varias veces a través del experimento.

f b) *Sensibilidad*. Ésta se refiere a la precisión con que la línea base (VD) puede "responder" a la VI. En otras palabras, se refiere a la forma en que los cambios de la VI son advertidos por un registro apropiado. Una línea base puede ser demasiado burda para advertir cambios que efectivamente ocurren en la conducta como consecuencia de cambios efectuados en la VI. Por ejemplo, si se pretendiera evaluar el efecto de un programa de modificación de conducta sobre la respuesta de ingerir bebidas alcohólicas y se decidiera usar como línea base (VD) el número de horas que una persona bebe diariamente, esta línea base podría indicar (falsamente) que el programa no estaría teniendo efecto. Podría suceder que la persona, como consecuencia del programa, hubiese disminuido ya sea el número de vasos con alcohol, el contenido alcohólico por vaso o ambos. Si éste fuera el caso, el programa (la VI) estaría teniendo un efecto sobre la conducta del sujeto; sin embargo, la línea base sería insensible a este efecto, a estos cambios que estarían ocurriendo y pasando inadvertidos al investigador.

c) *Control interno*. Éste se refiere a que la línea base "sea... tal que permita el control o la eliminación de procesos conductuales indeseables" (Sidman, 1960, pág. 321). Como ejemplos de dichos procesos indeseables que hacen confusa la descripción y evaluación de los datos, Sidman cita el caso en que una línea base es producto de

programas sobrelapados, el caso de líneas base con algún componente cíclico y el caso en que una respuesta íntimamente relacionada con la VD ocupa tiempo de registro de la línea base.

La figura 5 muestra una línea base que describe un proceso reversible. En este caso, se tenía un aparato en el cual el sujeto presionaba un botón, mientras un contador automático registraba cada presión y

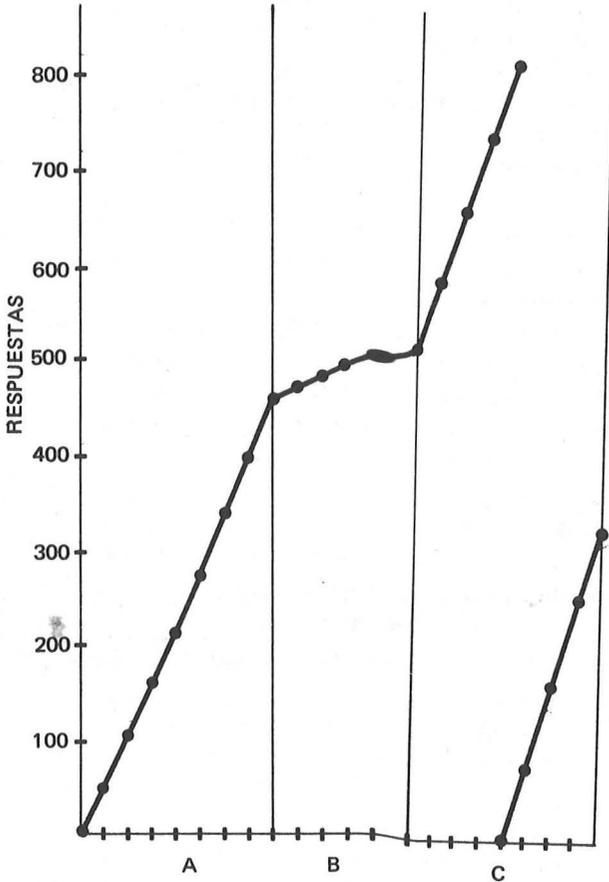


Figura 5. Registro acumulativo que muestra la tasa de una respuesta regida por una regla (fase B), comparada con la de línea base (A) y la recuperación opcional (C). (Tomado y adaptado de Moller, 1973.)

el experimentador registraba el tiempo. El sujeto era reforzado intermitentemente por la aparición de una luz con el letrero que indicaba "sí". Al final del programa recibía una indicación de cuántos "sies" (puntos) había obtenido. Aquí se puede notar que se tiene una situación de operante libre y que se está registrando la tasa de respuesta.

Este procedimiento produjo el segmento "A" de la línea base. A continuación se le indicó al sujeto que para obtener el máximo de puntos (número de "síes" multiplicados por una constante arbitraria), debería presionar el botón durante 10 segundos subjetivos, independientemente de que hubiera obtenido un "sí" o un "no" (ésta era la regla). Este procedimiento produjo el segmento "B" de la línea base. En la siguiente fase se le indicó al sujeto que el aparato estaba en un programa que le permitía emitir la conducta de la fase "A" o la de la fase "B" y que podría escoger una u otra (o ambas en sucesión) para responder. En este caso se produjo una reversión o recuperación del estado "A". Se puede notar que en un número bastante pequeño de respuestas se produjo una línea base estable y que ésta fue sensible a la manipulación experimental (Moller, 1973).

En algunas ocasiones se discute el papel que desempeña la línea base como sustituto del grupo control. Como se verá en la sección 4.6, la comparación entre ambos no es tan sencilla. Lo que sí se puede afirmar es que en la metodología del AEC, cada sujeto puede servir como su propio control cuando se tiene una línea base adecuada y manipulaciones experimentales precisas. Cuando se tiene una o más fases observacionales o de conducta estabilizada experimentalmente, éstas pueden servir como niveles de comparación dentro de un mismo individuo, a fin de evaluar los efectos de una o más VI. Se dice, entonces, que el sujeto sirve como su propio control. Cuando los efectos del orden de presentación de la VI son irrelevantes o cuando se está tratando de modificar una conducta específica sin buscar replicación alguna, no es necesario tener otros sujetos para propósitos de comparación. Estos propósitos se pueden satisfacer con la conducta del propio sujeto.

4.4. ¿UN SUJETO MEDIDO n VECES = n SUJETOS MEDIDOS UNA VEZ?

Una suposición frecuentemente utilizada en problemas de psicometría, teoría de la medición y escalamiento (véase Nunally, 1967, y Torgerson, 1967, por ejemplo) es la de que las personas son réplicas unas de las otras; es decir, una persona cualquiera es una "copia" o equivalente de otra persona cualquiera. De alguna manera, esta suposición (cuyo uso original era el de simplificar ciertos modelos de escalamiento) se ha aceptado tácitamente como un hecho y ha llevado al "principio" no formulado de equivalencia entre múltiples medidas de un solo sujeto y una medida de múltiples sujetos. Esto se ha extendido de una manera tal, que en algunas ocasiones (Estes 1967) se utilizan

los datos de un grupo de sujetos con el fin de estimar la probabilidad de respuesta de un organismo en particular (muchas veces ideal).

Esta suposición de replicabilidad de los sujetos prácticamente no se cumple en la naturaleza. La primera razón por la cual no se cumple es que los parámetros de respuesta varían considerablemente de individuo a individuo. Algunos responderán con poca frecuencia y otros con alta frecuencia; algunos con suma rapidez y otros con gran lentitud; sus umbrales perceptuales son diferentes, etc. . . . Ningún individuo sería capaz de responder con tales extremos de variabilidad. Por tanto, un individuo medido un gran número de veces (no importa qué tan grande) no puede ser equivalente a múltiples sujetos medidos una sola vez. Jamás podría ser tan versátil. La segunda razón por la cual no se cumple la suposición es que los datos obtenidos de muchos sujetos se suman (en promedios) para producir un "individuo ideal", el cual "debiera" ser equivalente al individuo real y al que eventualmente pudiera reemplazar. El "individuo ideal" no existe y no puede ser réplica de ningún sujeto real.

El principal error que se comete como consecuencia de creer que la suposición es un hecho, es el de realizar predicciones acerca del comportamiento del individuo real a partir del "individuo ideal". La razón es sencilla: si lo que se tiene como base de la predicción es un "individuo promedio" que responde idealmente en medio de dos extremos, entonces las predicciones describirán "las respuestas promedio" de dicho "individuo" en el futuro. Estas predicciones contendrán cierta cantidad de error (desde uno muy pequeño hasta uno enorme) cuando se comparen con el comportamiento real de sujetos particulares.

Por otra parte, el valor de producir dicho "individuo ideal" con un comportamiento ideal es el de describir sintéticamente a un gran número de sujetos y, además, el de estimar el valor más probable alrededor del cual se puede esperar una respuesta *de un grupo de sujetos* en el futuro.

Ejemplos de trabajos donde se consideran tanto los datos individuales como los promedios de un grupo de sujetos para hacer predicciones y poner a prueba modelos, se encuentran en el libro "ABC" (Atkinson, Bower y Crothers, 1965).

4.5. $N = 1$

Hay dos formas básicas de considerar los estudios con un solo organismo: *a*) aquella en la cual se trata de aplicar la metodología nomotética (en la que se trata de encontrar relaciones universales, y no particularidades) a la situación de $N = 1$; y *b*) aquella en la cual las

consideraciones se limitan al estudio "exhaustivo" de un solo individuo, sin interés alguno en la generalización.

En el primer caso, se debe considerar seriamente que es posible establecer relaciones funcionales con $N = 1$, siempre y cuando se ejerza el control experimental con todo rigor. El éxito rotundo del diseño reversible tanto en investigaciones con animales como con humanos apoya fuertemente la anterior afirmación. Por ejemplo, Boren (1966), al discutir el estudio de drogas con técnicas operantes, afirmó que un solo sujeto que sirva como su propio control y reciba todas las dosis de la droga, puede proporcionar una gran cantidad de información útil. "Ciertamente, bajo determinadas circunstancias y para ciertos propósitos, la información aportada por un solo sujeto puede ser... suficiente" (Boren 1966, pág. 551).

Una segunda consideración de importancia respecto a $N = 1$ tiene que ver con problemas de muestreo. El origen mismo del encabezado de esta sección sugiere que se le concede especial atención a N o al tamaño de la muestra. El tamaño de la muestra o el número de sujetos asignados a determinada condición experimental en los diseños tradicionales ha sido el objeto de innumerables escritos y discusiones respecto a la legalidad de las generalizaciones o a la representatividad de los datos. También se considera que afecta la validez externa de un experimento y la confiabilidad de los análisis de los datos. Otras consideraciones secundarias se relacionan a problemas de medición (error de medición y error de estimado) y a los efectos residuales de un tratamiento antes de iniciar otro (véase Gentile y colaboradores, 1972). Sin embargo, $N = 1$ es un nombre que confunde el hecho de que las generalizaciones se pueden efectuar con mayor seguridad mientras mayor sea el número de *observaciones* (o mediciones) implicadas, no importa (a este respecto) que éstas se obtengan a través de una observación en múltiples sujetos o mediante múltiples consideraciones en un solo sujeto. Incluso McNemar (1940), destacado estadístico contemporáneo, admite esta posibilidad.

Hay ciertos tipos de investigación conductual en los cuales la comparación de los datos de un individuo con otros individuos es especialmente irrelevante (por ejemplo, en el AEC o en psicofísica). A este respecto, Boren (1966) mencionó entre otras, dos razones para emplear solamente un sujeto; éstas son: *a*) los resultados obtenidos con este sujeto están de acuerdo con los datos obtenidos con otros en la misma área de investigación, y *b*) los resultados son consistentes con los de otros sujetos que el experimentador ha tratado con procedimientos similares. Duker (1965) presentó instancias en las cuales el uso de $N = 1$ se ve particularmente indicado por la bajísima frecuencia con que algunos fenómenos

ocurren. Entre ellos se encuentra el caso de ceguera unilateral al color, presentado por Graham y sus colaboradores, y el de insensibilidad congénita al dolor, señalado por Cohen y sus colaboradores. Otro ejemplo de esta clase es el del "mnemonista", estudiado por Luria (1968).

Además, se podría presentar abundante literatura donde se señala la toma de múltiples mediciones sobre el mismo sujeto. La mayoría de estos estudios pertenecen al área de modificación de conducta. Por ejemplo, en sus resúmenes de bibliografía selecta, Breger y Axelrod (1971) presentaron 70 estudios realizados entre 1965 y 1971, en los cuales se utilizó una "muestra" de $N = 1$.

También Duker señaló que, en ocasiones, el hallazgo y seguimiento de un solo caso ($N = 1$) que va en desacuerdo con una relación o ley aparentemente general, es suficiente para derrumbar dicha "ley". Gentile y sus colaboradores (1972), en un artículo reciente, sugieren que los datos obtenidos de un solo sujeto pueden tratarse con análisis estadísticos tradicionales (como AVAR) si tan sólo se abandona la suposición de que éstos son inadecuados para tratar casos de $N = 1$ o pequeñas muestras.

La otra clase de estudios es menos relevante en este punto, ya que simplemente se refiere al estudio clínico de casos particulares. En esta aproximación (ideográfica), el interés de seguir a un sujeto es principalmente el de una evaluación subjetiva de los efectos de algún "tratamiento clínico".

4.6. COMPARACIÓN ENTRE LÍNEA BASE Y GRUPO CONTROL

Esta comparación tiene como finalidad, por una parte, encontrar las semejanzas entre ambos conceptos y, por otra, hallar las diferencias que presentan. Como se verá en los párrafos siguientes, son muchas más estas últimas que las primeras.

El único punto en que parece haber algo en común entre la línea base y el grupo control es el que se refiere a su función como niveles de comparación: ambos permiten la evaluación del tratamiento experimental. Sin embargo, mientras que la función exclusiva del grupo control es la de permitir la evaluación de los efectos de la(s) VI, la línea base tiene múltiples funciones. Esto nos lleva a encontrar la primera diferencia entre ambos conceptos.

1. *Exclusividad.* Un grupo control tiene la función exclusiva de no recibir el tratamiento experimental o parte de éste para poder apreciar la magnitud del efecto de éste sobre la VD. Se puede decir, en este sentido, que el grupo control solamente asume el "estado de no

tratamiento". Por otro lado, la línea base es versátil; esto significa que, a veces, puede asumir estados puramente observacionales y, en otras ocasiones, diferentes estados manipulativos.

2. La diferencia principal entre la línea base y el grupo control es que *ambos obedecen a metodologías completamente diferentes, derivadas de modos muy divergentes de aproximar los problemas del comportamiento*. Mientras que en la aproximación tradicional de grupos es frecuente que los datos indiquen superficialmente cuáles son las relaciones entre dos variables o más (a través de análisis de tendencias, por ejemplo), en la aproximación del AEC las relaciones funcionales entre variables se pueden demostrar inequívocamente mediante el control experimental. El grupo control es producto de la primera aproximación y la línea base de la última.

3. *Suposiciones*. Mientras que el uso de uno o varios grupos control parte de ciertas suposiciones (o presunciones), el de la línea base no. Las principales suposiciones en las cuales se basa la formación de un grupo control son: *a*) por una parte, se considera válida la suposición de replicabilidad de los sujetos (véase la sección 4.4), es decir, que un sujeto es básicamente equivalente a otro; de tal manera, las diferencias individuales de respuesta no determinarán diferencias entre el comportamiento de sujetos controles y experimentales, y *b*) mediante procedimientos de aleatorización, apareamiento, etc., se logra una equivalencia entre las condiciones experimentales y controles, antes de la aplicación de la VI. La metodología de la línea base no parte de suposiciones semejantes.

4. *Sensibilidad de la VD al tratamiento*. Esta notable característica de la línea base no tiene contraparte dentro de la metodología tradicional. El investigador que emplea la metodología tradicional debe hacer estudios exploratorios para encontrar si una VI ejerce algún efecto sobre una VD, o, en el mejor de los casos, debe revisar la literatura y encontrar indicaciones de las relaciones entre sus variables de interés. Aun cuando consiguiera alguna confirmación de que la VI afecta a la VD, y a pesar de que sus propios datos le indicaran cuál es la probable forma de la relación funcional entre ellas, queda la duda de saber qué tan relevante o sensitiva es la VD. La VI pudiera estar afectando considerablemente a la conducta del sujeto y, aun así, la VD pudiera no estar reflejando este efecto. La relación entre la VI y la VD no es evidente durante el curso del experimento, y solamente se puede estimar hasta que éste ha terminado y los datos se han analizado. En la metodología de la línea base, por lo contrario, la relación entre VIs y VDs es evidente durante el transcurso del experimento (puesto que el registro de línea base casi siempre se genera

simultánea o inmediatamente después de la aplicación de la VI). La valuación de una o más VI se puede efectuar sin ambigüedades, debido a esta característica de sensibilidad de la línea base.

5. La siguiente diferencia consiste en que la línea base ofrece protección contra la variabilidad no controlada (o de muestreo), mientras que el grupo control no ofrece esta protección. La línea base evita la variabilidad de muestreo de dos maneras fundamentales: *a*) si la variabilidad es producida por las respuestas emitidas por el mismo sujeto, esto se evita mediante el logro de la estabilidad de la línea base (ya sea por medio de un gran número de observaciones o mediante un control efectivo sobre las variables que controlan a la línea base); *b*) si la variabilidad se produce como resultado de diferentes parámetros de respuesta de distintos organismos, entonces la metodología de la línea base permite su control a través de diversas técnicas de replicación (véase Sidman, 1960); *c*) la línea base tiene la ventaja de eliminar la variabilidad intersujetos, mientras que el grupo control no la puede eliminar (véase Sidman, 1960).

6. *Susceptibilidad al análisis estadístico.* La metodología tradicional (del grupo control) es altamente susceptible a una gran variedad de análisis estadísticos. Ya hemos mencionado los nexos inseparables que un gran número de autores establecen entre los diseños tradicionales y la estadística (Kirk, 1968, 1972; Edwards, 1968; Finney, 1960; Myers, 1972; Winer, 1962; Stanley, 1967; Plutchik, 1968, por ejemplo). Algunos autores como Sidman y Skinner, abiertamente identifican la aproximación tradicional con la estadística. Esta relación entre diseños tradicionales y estadística, no importa su grado, no debería extrañar a nadie. No debemos olvidar que tanto los diseños de grupo como ciertas técnicas estadísticas (como el análisis de varianza o el de regresión lineal múltiple) se desarrollaron paralelamente y que para poder aplicar algunos de estos análisis, se requiere que los datos (producto de la forma en que se agrupan los sujetos) cumplan ciertas condiciones, además de que las rápidas computadoras electrónicas permiten el desarrollo de métodos estadísticos sofisticados, especialmente adecuados para este tipo de diseño.

En suma, se podría afirmar que los diseños tradicionales son particularmente adecuados para el análisis estadístico de los datos que arrojan. Los diseños que siguen la metodología de la línea base, por otra parte, no dependen de la estadística, sino que generalmente no requieren de ella para el análisis de los datos, y en ocasiones los análisis estadísticos tradicionales no detectan cambios conductuales reales revelados por la línea base. Cabe mencionar aquí que están surgiendo nuevos tipos de análisis de los datos, especialmente adecuados para la meto-

dología de la línea base, como el análisis de tiempos interrespuesta (Skinner, Gott y Weiss, 1972) o la técnica *r* (Revushky, 1966) o el análisis de tiempos interreforzamiento (Fantino y Duncan, 1972).

7. *Múltiples reversiones.* Al usar la línea base como un medio para hacer de cada sujeto su propio control, es posible efectuar múltiples reversiones del tratamiento (vuelta a una línea base observacional o de un programa básico). Se puede investigar un gran número de tratamientos mediante esta técnica, siempre y cuando el cambio que se produce en la conducta no sea irreversible. Nada semejante se puede hacer cuando se trabaja con grupos control.

4.7. MODELO FIJO Y MODELO ALEATORIO EN EL DISEÑO DE GRUPOS

Los investigadores que emplean el AVAR suelen hacer una diferenciación respecto al tipo de modelo apropiado para la descripción de un diseño particular. En esta sección se analizarán brevemente dos de ellos: el modelo fijo y el modelo aleatorio. Se emplea el modelo fijo cuando los valores de la VI son escogidos por el investigador en una forma relativamente arbitraria, determinada por los propósitos de su experimento. Por ejemplo, en un experimento sobre condicionamiento clásico, el experimentador podría seleccionar los valores 10, 20, 30, 40 y 50 de la VI (número de ensayos) por creer que con ellos abarcaría un rango de valores lo suficientemente grande como para observar los efectos de esta VI sobre la VD (número de respuestas condicionadas). En este caso, se dice que los valores de la VI son fijos, ya que de todos los números enteros entre 0 y 50, solamente cinco de ellos pudieron estudiarse, dependiendo de la selección "arbitraria" del investigador. En términos más formales, se diría que no todos los valores contenidos en el rango de la VI tuvieron la misma probabilidad de ser seleccionados.

Por otra parte, se usa el modelo aleatorio cuando cada valor en el rango de la VI tiene la misma probabilidad de ser seleccionado para una manipulación experimental. Cuando el experimentador emplea este modelo, lo único que selecciona personalmente es el rango de la VI. Los valores de la VI se seleccionan al azar (de acuerdo a una tabla de números aleatorios, por ejemplo), cumpliéndose así el requisito de que cada valor tenga la misma probabilidad de ocurrir.

Las implicaciones del uso de uno u otro modelo son importantes en dos aspectos: *a*) para seleccionar adecuadamente los estimados de varianza necesarios en los cómputos del AVAR, y *b*) en los límites

que se le imponen a la generalización. El primer aspecto no lo analizaremos aquí, por ser básicamente una cuestión estadística. El segundo es de interés más general, particularmente relevante al diseño experimental en el cual participan diversos grupos de sujetos.

La limitación básica que implica el uso del modelo fijo es que el investigador solamente puede hacer afirmaciones acerca de los valores de la VI que realmente manejó, y no respecto de ningún otro; no deberá interpolar los datos (referirse a valores intermedios no manipulados), ni tampoco extrapolar (hacer inferencias acerca de valores fuera del rango de la VI). Si, por ejemplo, fuese a graficar los datos del caso de condicionamiento clásico arriba mencionado, solamente debería dibujar un punto por cada valor de la VI, y bajo ninguna circunstancia debería unir los cinco puntos; si lo hiciera, estaría excediendo los límites de generalización que el modelo le permite, o sugiriendo, falsamente, que manipuló los valores intermedios (digamos, entre 10 y 20 en nuestro ejemplo).

Por otra parte, las implicaciones son diferentes cuando se utiliza el procedimiento correspondiente al modelo aleatorio. En este caso no existe la limitación respecto a la interpolación de los datos (unión de los puntos con una línea), ya que cada valor en el rango de la VI tiene la misma probabilidad de ocurrir. Cuando al usar este modelo, se generaliza a los valores no manipulados dentro del rango de la VI, se considera que los valores muestreados representan adecuadamente el resto de los valores contenidos en el rango. Por tanto, este modelo permite al investigador generalizar sus conclusiones a todos los valores dentro del rango en cuestión, sin importar que no los haya manipulado realmente (interpolación). Frecuentemente, hay la tendencia a ignorar la distinción entre el modelo fijo y el modelo aleatorio. Cuando esto sucede, el investigador pudo haber incurrido en dos errores: *a*) haber utilizado un procedimiento computacional equivocado, que hubiera producido datos engañosos, y *b*) haber sobregeneralizado el presentar y analizar sus resultados.

Estas implicaciones respecto a la generalización producidas por el uso de uno u otro modelo se pueden extender sin dificultad a todo diseño experimental de grupos e, incluso, a diseños en los cuales se mide repetidamente al mismo sujeto.

5

diseños experimentales. generalidades y clasificación

En nuestra búsqueda de nueva información debemos estar preparados, en cualquier momento, a cambiar nuestra concepción de lo que es un diseño experimental deseable... El diseño experimental adecuado no puede legislarse por principios lógicos ni por principios empíricos.

MURRAY SIDMAN

EL EMPLEO DE diseños experimentales en la ciencia de la conducta es visto por diversos autores en formas no sólo diferentes sino, en ocasiones, conflictivas. Por una parte, está la posición de Skinner (Evans 1968, Skinner, 1969), quien niega valor alguno a la actividad de diseñar experimentos; por otra, está la posición estadística tradicional, que considera a la experimentación como un mero elemento para evaluar ciertas hipótesis estadísticas. Por ejemplo, Fisher (1966) considera que "...se puede decir que todo experimento existe, solamente para dar a los hechos una oportunidad de desaprobar la hipótesis nula", aunque en otra parte afirma que el procedimiento estadístico y el diseño experimental son dos aspectos diferentes de la misma totalidad, y la tendencia general en su libro es la de relacionar ambos aspectos; y Dixon y Mossey (1969) afirman que "el método de escoger una muestra se designa como el *diseño del experimento*".

En medio de estos dos extremos hay toda clase de énfasis sobre diferentes aspectos del diseño experimental; por ejemplo, está la posición autocorrectiva de Sidman, descrita en la cita inicial de este capítulo. También está la posición de Brunswik (en Bachrach, 1972) respecto al logro de un diseño representativo, o sea, un compromiso entre maximizar el control experimental y minimizar la artificialidad. Asimismo, está la posición futurista de Baker (1967), quien considera que la tradición fisheriana de diseños experimentales será incapaz de ajustarse a la investigación de gran escala del futuro inmediato. Este autor indica que los diseños experimentales tradicionales fueron elaborados, principalmente, para resolver problemas de la agricultura (o del investigador con pocos medios) y que su principal característica era la de que se aplicaban a experimentos aislados.

Ahora bien, el estado presente del desarrollo científico y tecnológico no solamente implica el uso de gigantescas computadoras y de procedimientos multivariados, sino, según Baker, el planeamiento de complejas series de experimentos, en las cuales el resultado de un experimento debería determinar la forma del siguiente, todo esto dentro de una estrategia flexible y con una meta determinada de antemano. Baker llama a esto "la conceptualización del nuevo diseño experimental", y propone que métodos como la investigación de operaciones o el diseño de superficie de respuesta pueden servir de base para desarrollar dicho sistema de diseño experimental.

Recientemente, Herman y Arbeit (1973) informaron de una serie de experimentos, en los cuales se seguía un diagrama de flujo, indicando qué procedimientos experimentales se debían seguir de acuerdo al comportamiento del sujeto (un delfín). Este tipo de planificación experimental parece corresponder a las expectativas de Baker, y será una parte indispensable del diseño experimental a medida que aumente la complejidad de la conducta a estudiar y de la instrumentación y técnicas implicadas.

Cualquiera que sea la posición respecto al uso de diseños experimentales o al empleo de estadística o limitación de una clase de diseños, hay un acuerdo general en cuanto a qué es lo que constituye un diseño. El aspecto en el que hay un consenso más o menos general lo constituye el concepto más simple de diseño experimental. De acuerdo a este concepto, un diseño experimental, en la ciencia de la conducta, es el arreglo de las condiciones experimentales —valores de la(s) VI— por las cuales habrá(n) de someterse uno o más organismos (unidades experimentales), de acuerdo a la forma de la pregunta de investigación.

Hay otros aspectos más específicos en los cuales el acuerdo ya no es tan grande. Veamos lo que dicen diferentes autores. Para Kirk

(1972), hay cinco actividades interrelacionadas que caracterizan al diseño experimental: *a*) la formulación de hipótesis estadísticas y el plan para obtener los datos que permitan evaluar dichas hipótesis; *b*) el establecimiento de reglas de decisión para probar las hipótesis estadísticas; *c*) la colección de los datos de acuerdo al plan; *d*) el análisis de los datos, y *e*) la toma de decisiones respecto a las hipótesis y a la formulación de inferencias inductivas concernientes al grado de "veracidad" de las hipótesis científicas.

Para Finney (1960), el diseño de un experimento está constituido por *a*) la serie de tratamientos seleccionados para hacer comparaciones; *b*) la especificación de las unidades a las cuales se aplicarán los tratamientos; *c*) las reglas por las cuales se asignarán los tratamientos a las unidades experimentales, y *d*) la especificación de las medidas a ser tomadas de cada unidad (VD).

Para Plutchik (1968), el diseño experimental se define en términos de: *a*) las formas de arreglo de condiciones experimentales indicadas por la pregunta de investigación, y *b*) los métodos de control para minimizar o reducir el error al establecer relaciones causales entre las variables implicadas. Para Lathrop (1969), lo que determina un diseño es la selección de un problema. Para D'Amato (1970), las consideraciones respecto a los análisis estadísticos factibles son determinantes en la selección de un diseño. Finney concuerda en este aspecto y afirma que: "...el biólogo implicado en cualquier evaluación cuantitativa debe usar métodos estadísticos, no importa si les da ese nombre o no" (Finney, 1960, pág. 2). Edwards (1968) no da definición alguna, sino que simplemente procede a crear "diseños" que corresponden al uso de algunas técnicas derivadas de la teoría de la probabilidad, para hacer la evaluación de un resultado observado. Johnson y Solso (1971) hacen énfasis en la función del diseño experimental, para lograr el control de la VI y el de las VE. Sheridan (1971) considera que el diseño proporciona al experimentador: *a*) un bosquejo para disponer el experimento con los controles apropiados, y *b*) un lenguaje para la comunicación con sus colegas. Para Myers (1972), el diseño experimental es el plan para llevar a cabo el experimento, y hace énfasis en tres consideraciones básicas para seleccionar el diseño: *a*) la información buscada: diferentes diseños proporcionan distinta información, y se pueden hacer simplificaciones en un diseño, a expensas de cierta información; *b*) el modelo; ésta es una consideración que concierne a los diseños tradicionales, en que se debe tener en cuenta el modelo estadístico relacionado con el diseño que se va a utilizar, y *c*) las restricciones prácticas, como disponibilidad de sujetos, equipo, presupuesto, etc.

Anderson (1969) considera tres tipos principales de diseño experimental, a saber: *a*) diseños de sujetos independientes, en los cuales se asignan diferentes sujetos en una forma aleatoria a las distintas condiciones experimentales; *b*) diseños de sujetos apareados, en los cuales los sujetos son apareados de acuerdo a una "variable pertinente", y *c*) diseños de sujetos idénticos, en los cuales los mismos sujetos son sometidos a todas las diferentes condiciones experimentales. Anderson considera otros aspectos marginales al diseño; entre ellos, la manipulación de la VI, medición de la VD, procedimiento, tipo de análisis de los datos que se va a hacer, y una anticipación de las diferentes formas que pudieran asumir los resultados.

McGuigan (1972) tiende a la misma aproximación. Winer (1962) diferencia claramente las funciones de "diseñador experimental" y de "experimentador". Para él, la primera se refiere a las actividades que se requieren para hacer el plan o estructura del experimento; el sujeto que mejor podrá desarrollar dichas actividades será aquel que: *a*) esté más familiarizado con el material experimental; *b*) esté más familiarizado con los diseños alternativos, y *c*) sea el más capaz de evaluar los diseños factibles, en términos de sus ventajas y desventajas relativas. La segunda función se refiere a la de *ejecutor* del experimento, es decir, la persona que se enfrenta a los sujetos, manipula los instrumentos de medición o estimulación, etc. En un buen número de casos, un mismo individuo podría asumir ambas funciones, pero esto no es indispensable.

Kemphorne (1952), por otra parte, considera al diseño experimental como "...esencialmente el patrón de las observaciones a ser colectadas". Con el propósito de comprobar ciertas hipótesis, divide a los experimentos en dos clases: *a*) *absolutos*, y *b*) *comparativos*. A la primera clase pertenecen aquellos en que se determinan ciertos parámetros (por ejemplo, la carga eléctrica de un electrón, los umbrales sensitivos o el alcance —espacio-amplitud— de la memoria inmediata). A la segunda pertenecen, afortunada o desafortunadamente (véase Simon, 1969), la mayoría de los experimentos en psicología y ciencias afines. Kemphorne afirma que en estos experimentos, el análisis de un solo tratamiento experimental es de poco interés por sí solo. Son las inferencias derivadas de las comparaciones las que tienden a generar nueva investigación y las que proporcionan la mayoría de los datos conductuales.

La *Enciclopedia de psicología*, de Dushkin (1973), define al diseño experimental como "una serie de procedimientos precisamente planeados para determinar las relaciones entre variables, lo que permite la predicción de la conducta de los organismos" (pág. 93).

De los autores consultados, una gran mayoría (Finney, Sheridan, Kirk, Myers, Anderson, Johnson y Solso, Plutchik, Anderson, Edwards, McGuigan, Winer, D'Amato, Lathrop, Sidman, etc.), consideran, explícita o implícitamente, al arreglo u organización de condiciones, a las reglas para la asignación de unidades experimentales o tratamientos (o viceversa) y a sus relaciones con la pregunta de investigación, como el núcleo del diseño experimental. No está de más subrayar que sólo en este núcleo se encuentra tal acuerdo entre ellos. Otros aspectos, enfatizados por unos y no por otros como características del diseño experimental, son:

- a) La especificación de la VD (Plutchik, Anderson, McGuigan y Finney);
- b) los métodos de control (manipulación) de las VI y las VD (Johnson y Solso, Plutchik, Anderson y McGuigan);
- c) la comunicación entre colegas (Lathrop);
- d) el procedimiento experimental (Anderson y McGuigan);
- e) la selección de la muestra (Dixon y Massey); y
- f) alguna variante de este esquema, formulación de hipótesis, estadísticas, reglas de decisión, colección de datos, análisis de los datos, inferencias estadísticas y lógicas (Kirk, Fisher, Dixon y Massey, McGuigan, Edwards, Cochran y Cox, Winer y Linqvist).

Los primeros cuatro puntos simplemente revelan un desacuerdo en cuanto a la extensión del concepto de diseño experimental con respecto al núcleo antes mencionado. Los dos últimos aspectos son críticos, ya que advocan que la estadística desempeña un papel fundamental en el diseño experimental, tesis básicamente en conflicto con la de este libro, a saber: que *es posible desarrollar y usar diseños experimentales independientemente de la estadística*.

Algunas de las fallas más graves de la aproximación estadística fisheriana a la experimentación conductual son analizadas minuciosamente por Bakan (1966). Su análisis se centra en las llamadas "pruebas de significancia", basadas en la formación de la hipótesis nula. Su primer punto es el de que la hipótesis nula representa un evento que tiene una probabilidad infinitamente pequeña de ocurrir en la naturaleza, cuyo rechazo es factible siempre que se tenga un número relativamente grande de datos. El segundo punto relevante a este análisis es su crítica a la pretensión (atribuida a Fisher) de automatizar el proceso de la inducción científica.

El uso de las reglas de la inferencia estadística, formulado primero por Fisher y elaborado posteriormente por sus seguidores, ha llegado,

en algunos medios, a ser considerado como sustituto de la inferencia estadística. Bakan afirma que la responsabilidad de hacer inferencias inductivas, de estar equivocado, "...se cambiaron de los hombros del investigador y se pusieron en las pruebas de significancia" (Bakan, 1966, pág. 430). Finalmente, el tercer punto es que en la aproximación estadística de Fisher hay una confusión de dos niveles de inferencia.

El primero es el siguiente: de acuerdo con el modelo de la estadística inferencial, es posible hacer afirmaciones acerca de una población a partir de los datos de una muestra obtenida de aquélla. Éste es un procedimiento perfectamente válido. Se hace una inferencia muestra-población.

El segundo nivel, que se pretende hacer contingente al primero (automatizando así la inducción científica), no está indicado en lo absoluto por el modelo. No hay nada en el modelo que valide la identificación de la inferencia estadística con la inferencia científica.

En suma, Bakan critica acerbamente el que se intente identificar al proceso de inducción científica con el de la inferencia estadística, que implica hacer pruebas de significancia y evaluar hipótesis nulas de escasa relevancia científica.

Otra falla grave, que es posible notar simplemente revisando el índice de un libro sobre diseño experimental, consiste en la errónea identificación entre los diversos diseños experimentales y los modelos estadísticos fundamentales. Se ha informado de experimentos en los que se emplea un diseño "2 × 2 AVAR", o se mencionan diseños AVAR en general o de control estadístico. Todos estos puntos indican cómo, exagerando la aproximación de Fisher, se sustituye al diseño experimental por su contraparte estadística y se intenta complementar al control experimental. Este fuerte énfasis en la estadística hace que, a los ojos de autores, editores, estudiantes e investigadores, el diseño experimental parezca irremediablemente atado a aquélla. Además, con tal de no violar las suposiciones de los modelos estadísticos, más de un investigador ha tenido que cambiar su experimento, aumentando el número de sujetos o eliminando a aquellos que obtuvieron puntuaciones demasiado "desviadas", o de alguna otra manera.*

El hecho es que se llevaban a cabo experimentos, antes de que surgiera la estadística que los dictaminara (o tratara sus datos, en el mejor de los casos). El diseño reversible se empleó en otras ciencias, antes que en la psicología, y no fue necesario recurrir a la inferencia

* Afortunadamente, ya algunos autores se han adjudicado la tarea de, positivamente, violar las suposiciones del análisis de varianza (Young y Veldman, 1963), así como algunas del análisis de covarianza (Calkins, 1971), encontrando que cuando N no es demasiado pequeña, las violaciones tienen consecuencias insignificantes.

estadística para poder evaluar los efectos de la VI. La estadística simplemente ayudó a hacer análisis más complejos y precisos de cierto tipo de datos. Estos análisis debieran ayudar al investigador a apreciar el efecto de la VI, especialmente en aquellos casos en que éste es dudoso. La pregunta relevante parece ser, entonces, ¿qué tanto queda del diseño experimental cuando no se usa en absoluto la estadística? En los siguientes capítulos trataremos de contestar esta pregunta específicamente, pero, por el momento, adelantamos una respuesta general y sugerente: lo que queda es una organización lógica de las condiciones experimentales, lo suficientemente sensible como para contestar la pregunta de investigación. Debe entenderse claramente que no se está propugnando aquí porque se elimine la cuantificación del diseño experimental (lo cual sería absurdo en este contexto), ni siquiera porque se elimine la estadística. Simplemente, se pretende reducir la influencia de la estadística hasta su papel adecuado: el de un auxiliar para el análisis de datos. Veamos ahora la descripción y justificación del sistema clasificatorio que vamos a utilizar.

La clasificación que usaremos es doble. Por un lado, se tiene en cuenta el número de VIs que se manipulan en un diseño y la cantidad de valores que cada una de ellas asume. La justificación de esta parte de la clasificación es bastante simple y directa. Recordemos que el arreglo de las condiciones que constituyen un diseño depende de los valores de la(s) VI(s). Si los diseños se forman como una consecuencia de las VIs que participan en una pregunta de investigación, entonces lo más acertado es describir a los diseños resultantes de acuerdo a las VIs que los determinaron. Por otro lado, se tiene en cuenta el tipo de metodología que determina el uso (o desarrollo) de los diseños. Consideraremos que las dos metodologías dominantes en la ciencia de la conducta son la experimentación tradicional con grupos de sujetos y el AEC. La justificación para usar esta parte de la clasificación se basa en el reconocimiento de dos aproximaciones experimentales fundamentalmente diferentes, que emplean distintos diseños experimentales y que dominan la literatura conductual contemporánea. En este libro se les coloca una al lado de la otra, a fin de facilitar su comparación.

5.1. CLASIFICACIÓN DE ACUERDO CON LA(S) VARIABLE(S) INDEPENDIENTE(S)

La clasificación en base a las VIs es simple, pues solamente contiene dos clases:

1. Los *diseños univariantes*; en este tipo de diseño se maneja sólo una VI, irrespectivamente, de la metodología empleada. Esta clase se divide, a su vez, en dos subclases, de acuerdo con el número de condiciones contenidas en el diseño; éstas son: *a)* diseños bicondicionales, y *b)* diseños multicondicionales.
2. Los *diseños multivariantes*; en este tipo de diseños interviene más de una VI (teóricamente, sin un límite superior) también independientemente de la metodología a la que pertenezcan. Esta clase de diseños se divide en las siguientes cuatro subclases: *a)* diseños reversibles multivariantes, *b)* diseños factoriales, *c)* diseños jerárquicos, y *d)* diseños incompletos.

Esta es parte de la clasificación básica que usaremos en este libro, y los siguientes capítulos se han elaborado de acuerdo a ella. A los lectores interesados en otros sistemas clasificatorios de los diseños experimentales se les recomienda consultar la obra *Diseños experimentales* (Cochran y Cox, 1971), el trabajo de Kirk (1972) y el de Winer (1962). La otra parte de la clasificación que se emplea en este libro está relacionada con el tipo de metodología concerniente a los diseños; la analizaremos a continuación.

5.2. CLASIFICACIÓN POR TIPO DE METODOLOGÍA. DISEÑOS CONDUCTUALES Y DISEÑOS TRADICIONALES

Pudiera parecer impropio denominar a los diseños de tipo reversible “diseños conductuales”, ya que este tipo de diseño se ha empleado en áreas como la psicofísica, dentro de la psicología, o en ciencias como la fisiología o la física, fuera de ella. La razón para identificar este tipo de diseño con el AEC consiste en que el mismo es dominante en la investigación y aplicación del AEC. La razón para analizar este tipo de diseño es la creciente relevancia de la metodología del AEC en la psicología y en disciplinas afines.

Un diseño reversible es aquel en el cual se realizan diferentes series de observaciones sobre la misma unidad experimental, teniendo siempre la posibilidad de “regresar” a una etapa experimental anterior. Como veremos más adelante, hay un diseño de tipo reversible en el cual solamente se pasa a la unidad experimental por dos etapas, sin efectuar realmente la reversión. A esto podríamos considerarlo como un diseño reversible “incompleto”, aun cuando algún autor más estricto no lo clasificaría como reversible.

En el resto del libro adoptaremos el acuerdo de identificar diferentes tipos de diseños reversibles como diseños conductuales y los analizaremos esencialmente con ejemplos de dicha metodología. Por tanto, la primera categoría de esta clasificación se halla constituida por "los diseños conductuales".

El segundo tipo de diseños de esta clasificación es el de los diseños tradicionales. Por éstos se entienden *todos aquellos diseños de grupo desarrollados en la tradición fisheriana-estadística*. A esta clasificación pertenecen los diseños que generalmente se presentan en los textos sobre diseño experimental en la ciencia de la conducta, la biología y la agricultura. Solamente serán estudiados los más representativos, haciéndose énfasis en su estructura lógica, sus puntos fundamentales y sus limitaciones o desventajas relativas.

6

diseños univariabes bicondicionales

LOS DISEÑOS QUE aportan el mínimo de información necesario para hacer una comparación formal son aquellos en que se maneja una VI y que tienen dos condiciones experimentales. Los principales usos de esta clase de diseño son el de producir información de tipo exploratorio y el de servir para los propósitos de un gran número de investigaciones aplicadas. En el primer caso, un diseño de este tipo puede servir para indicar en una forma tentativa si hay algún efecto de la VI sobre la VD y cuál es su dirección. En el segundo caso, puede servir la función de evaluar con cierto grado de certeza la efectividad de un tratamiento para la solución de un problema dado.

I. CLASE CONDUCTUAL

6.1. DISEÑOS A-B

Aun cuando éste no es un diseño reversible, comparte con dicho tipo ciertas características; tan es así, que uno se ve tentado a llamarlo “diseño potencialmente reversible” o “diseño reversible incompleto”.*

* Algunos autores lo refieren como “diseño de medidas repetidas AB”.

En los experimentos con este diseño, el mismo organismo o grupo de organismos participa en ambas etapas del diseño. Generalmente, se usa con propiedad la metodología de la línea base y, por lo común, el análisis de los datos no es de tipo estadístico. Se pueden encontrar en la literatura excepciones a los dos últimos puntos, pero no al primero.

Se considera diseño A-B cualquier arreglo de condiciones que: *a*) tenga claramente definidas dos etapas experimentales, en una de las cuales se aplica la VI de interés; *b*) la etapa B suceda temporalmente a la etapa A; *c*) proporcione una línea base para evaluar los efectos de la VI (éste es un punto sumamente importante, ya que descalifica a diseños del tipo pretest-postest); *d*) el mismo organismo o grupo de organismos pase por las dos etapas, y *e*) caracterice claramente la conducta registrada en la línea base de la etapa A. Dado que ésta servirá como nivel de comparación, es necesario que la conducta sea estable y que corresponda a un arreglo de contingencias conocido. Es común referirse a esta etapa como *observacional*, pero éste no es necesariamente el caso, ya que se tienen líneas base en las cuales el experimentador las determina completamente (por ejemplo, cuando se somete al organismo a un programa de intervalo variable).

Veamos algunos casos, tanto de investigación básica como de investigación aplicada, en los que se utilizó este diseño. Dews (en Honig, 1966) comparó el registro acumulativo de una paloma que discriminaba colores bajo condiciones normales (programa de IV), logrando que solamente respondiera en presencia de D con el registro acumulativo producido bajo la influencia de algunos mg de pentobarbital. Se observó la interrupción del proceso de discriminación previamente producido. Skinner (1938) reportó un experimento en el cual primero se sometió a un animal a un programa de "recondicionamiento periódico" de 5 minutos (fase A) y a continuación a un programa de razón fija (fase B), obteniéndose un marcado incremento de la tasa de respuesta. Koegel y Covert (1972) estudiaron la adquisición de conducta discriminativa en niños "autistas". En su primer experimento registraron el porcentaje de discriminación correcta y el de conducta autoestimulativa durante la línea base (A). Entonces trataron de encontrar si la conducta autoestimulativa tenía alguna relación con la discriminación. Para ello, procedieron a suprimir la conducta autoestimulativa por medio de indicaciones verbales (un fuerte "no") o mediante manipulación física brusca (fase B). La supresión de la autoestimulación tuvo el efecto de aumentar las discriminaciones correctas, hasta lograr su adquisición. La figura 6 muestra los resultados de estos autores, en una replicación de las fases A-B (sólo aparece uno de sus sujetos).

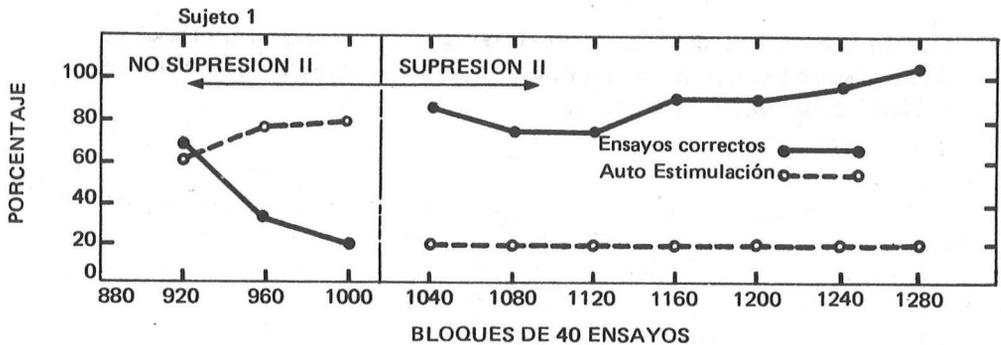


Figura 6. Porcentaje de ocurrencias de autoestimulación y de ensayos correctos para el sujeto 1 durante la segunda condición de "no supresión". A este sujeto se le aplicó una segunda condición de "supresión" de la autoestimulación. (Tomada y adaptada de Koegel y Covert, *Journal of Applied Behavior Analysis*, vol. 5, núm. 4, verano de 1972, págs. 381-387. Copyright 1972 de la Society for the Experimental Analysis of Behavior, Inc.)

Dunham (1972) estudió los efectos del castigo (choque eléctrico) sobre respuestas inicialmente no castigadas. Durante la fase A (línea base observacional) las ratas tuvieron libre acceso a un tubo-bebadero y a una rueda de actividad; el experimentador simplemente registró la frecuencia con que ocurrían estas respuestas y derivó su probabilidad. En la fase B incluyó la VI de castigo: aplicaba un choque eléctrico inmediatamente después de cada respuesta de chupar en el tubo-bebadero. Observó un gran aumento en la respuesta de correr y un notable decremento en la conducta de beber. En un experimento subsecuente empleó el mismo diseño, siendo idéntica la fase A a la descrita; pero en la fase B, la manipulación experimental consistió en la eliminación del tubo-bebadero o en frenar mecánicamente la rueda de actividad. Se observó que al impedir la emisión de una de las dos respuestas, la otra aumentaba considerablemente.

Aunque, estrictamente hablando, Schneider y Neuringer (1972) no utilizaron un diseño A-B, analizaremos brevemente su investigación por ser especialmente ilustrativa. Estos investigadores condicionaron cinco palomas a responder bajo un programa de operante restringida (intervalo fijo en ensayos discretos), produciendo dos estados de conducta que se aproximan considerablemente a dos estados fijos. Aun cuando sus registros acumulativos son similares a los que se producen bajo un programa de intervalo fijo, ellos reagruparon sus datos de acuerdo a los estados fijos "antes del punto de corte" o "después del punto de corte". Sus gráficas ilustran claramente el cambio conductual obtenido al pasar de un estado al otro. Un diseño A-B en el que se

lograse un control efectivo sobre la conducta estudiada, debería producir resultados tan impresionantes como los obtenidos por Schneider y Neuringer (véase la fig. 7).

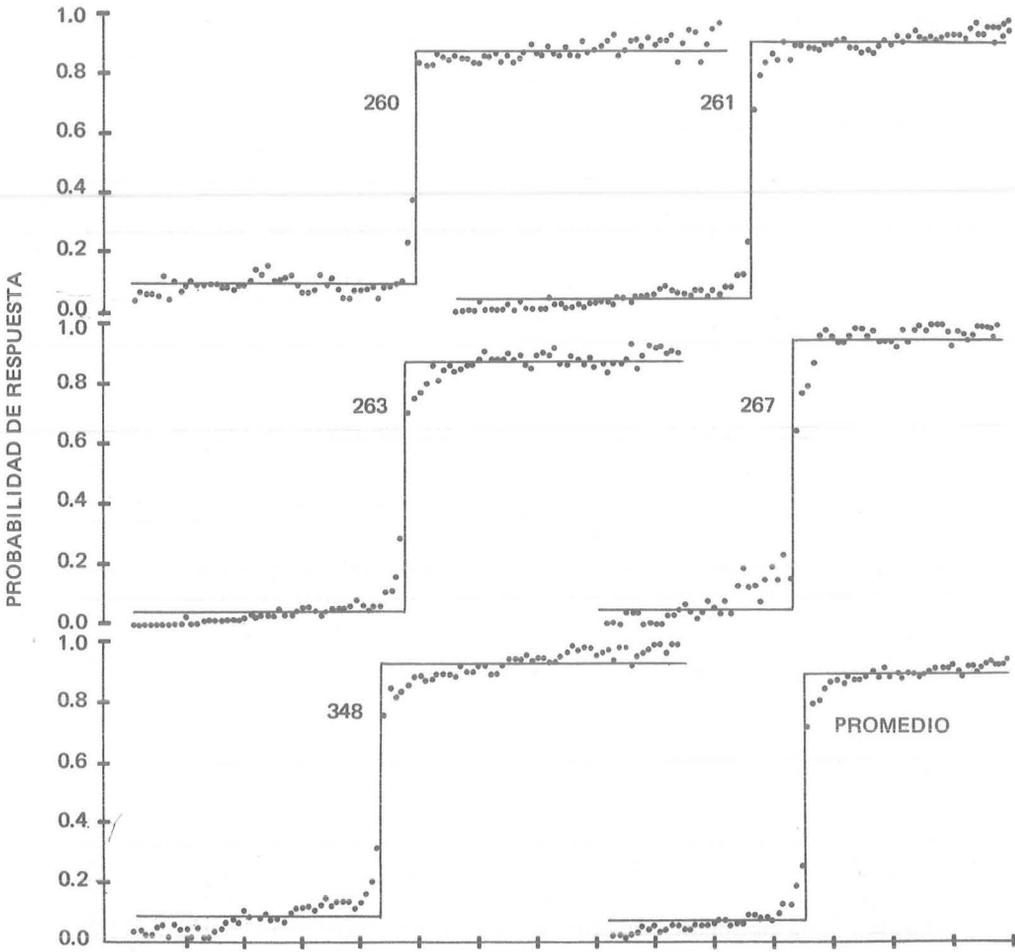


Figura 7. Probabilidad de respuesta antes y después del punto de corte, para cinco pájaros en IF 252 seg de ensayos discretos. La figura denominada "promedio" indica la media de esos cinco pájaros. La función sólida de escalón representa la conducta de dos estados perfectos. Cada unidad a lo largo de la abscisa corresponde a 40 segundos (diez ensayos). Cada punto representa un ensayo. (Tomado de Schneider y Neuringer, *Journal of the Experimental Analysis of Behavior*, vol. 18, núm. 2, 1972, págs. 187-199. Copyright 1972 por la Society for the Experimental Analysis of Behavior, Inc.)

Otro ejemplo en el que parte de una investigación se puede considerar como correspondiente a un diseño A-B fue presentado por de Villiers y Millenson (1972). Después de entrenar tres ratas a accionar cualquiera de dos palancas bajo dos programas independientes de intervalos aleatorios concurrentes, incluyeron periodos de "peligro" en los cuales una luz servía como señal para indicar la presentación de un choque eléctrico inevitable. Cuando se eliminaron los choques, las respuestas aumentaron considerablemente (correspondiendo al nivel alcanzado con anterioridad a la inclusión de los periodos de "peligro"). En una de sus figuras, Villiers y sus colaboradores muestran el cambio notable que se produce al pasar de los periodos de "peligro" (A), a la extinción de la supresión condicionada (B) (véase la figura 8).

Por otra parte, Christophersen y sus colaboradores (1972) instruyeron a padres de familia a aplicar un sistema de economía de fichas (puntos) en sus hogares, a fin de modificar la conducta de sus hijos. Las conductas a ser modificadas (tanto las que ganaban puntos como las que los perdían) fueron definidas adecuadamente. También se especificó cuáles eran las conductas terminales deseadas, los reforzadores que se canjearían por puntos y los valores específicos de reforzadores y conductas. La figura 9 muestra los niveles de línea base (A) de diferentes conductas emitidos por tres niños, comparados con los niveles alcanzados cuando se incluyó el sistema de puntos (B); el incremento en todas las conductas fue notable. La figura 10 muestra el efecto que tuvieron las "multas" (pérdidas de puntos) sobre ciertas respuestas consideradas como "indeseables" por los padres; de nuevo, el cambio de la fase A a la B fue notorio. Finalmente, la figura 11 presenta un análisis de línea base múltiple de un solo sujeto: en la fase A no se multó al niño, y en la fase B se impuso una multa de 10 puntos. El efecto del tratamiento sobre las tres respuestas fue evidente.

Bandura y McDonald (en Bandura, 1969) utilizaron el diseño A-B, en cuya primera fase registraron el número de "juicios morales objetivos" en niños que presentaban una forma predominantemente subjetiva de emitir "juicios morales". Después (durante la fase B) algunos de estos niños tuvieron la oportunidad de observar adultos que emitían juicios morales objetivos, y eran reforzados cada vez que expresaban un juicio de este tipo. Los demás niños no tuvieron la oportunidad de emular un modelo, pero también fueron reforzados cada vez que emitían un juicio moral objetivo durante la fase B del experimento. Se observó que los niños que sólo recibieron reforzamiento por emitir juicios objetivos, prácticamente no aumentaron la frecuencia de esta clase de respuestas. Sin embargo, los niños que pudieron imitar un modelo emitieron más respuestas durante el tratamiento (fase B).

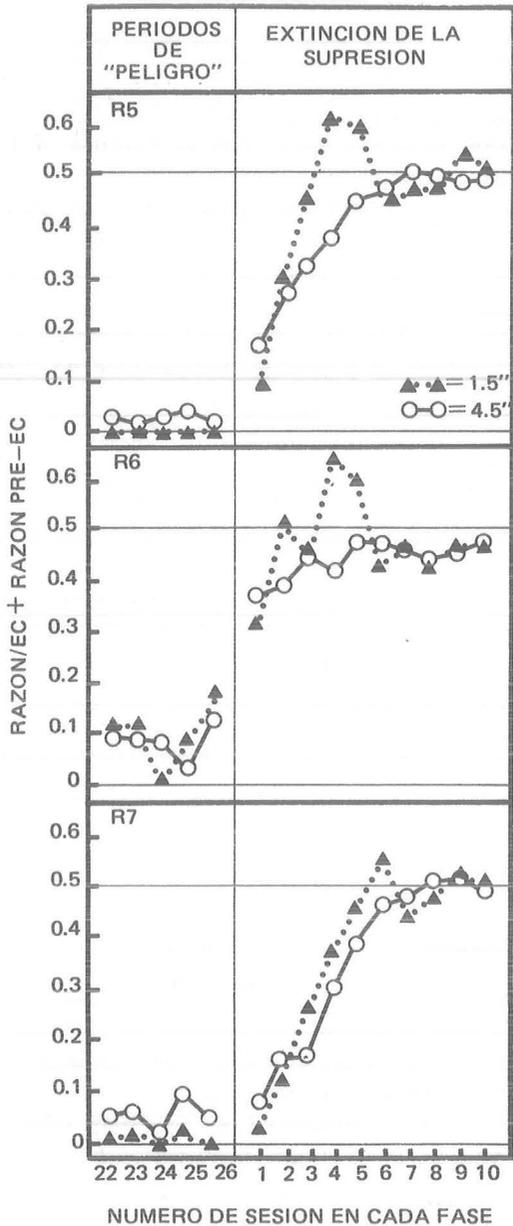


Figura 8. Razones de supresión para las cinco últimas sesiones de periodos de peligro de presentación de choque y para las subsecuentes diez sesiones de extinción. Los triángulos sólidos representan datos basados en las respuestas asociadas con una duración del reforzador de 1.5 segundos. Los círculos abiertos representan datos basados en respuestas a la palanca, asociados con una duración del reforzador de 4.5 segundos. (Tomado de De Villiers y Millenson, *Journal of the Experimental Analysis of Behavior*, vol. 18, núm. 2, 1972, págs. 287-294. Copyright 1972 por la Society for the Experimental Analysis of Behavior, Inc.)

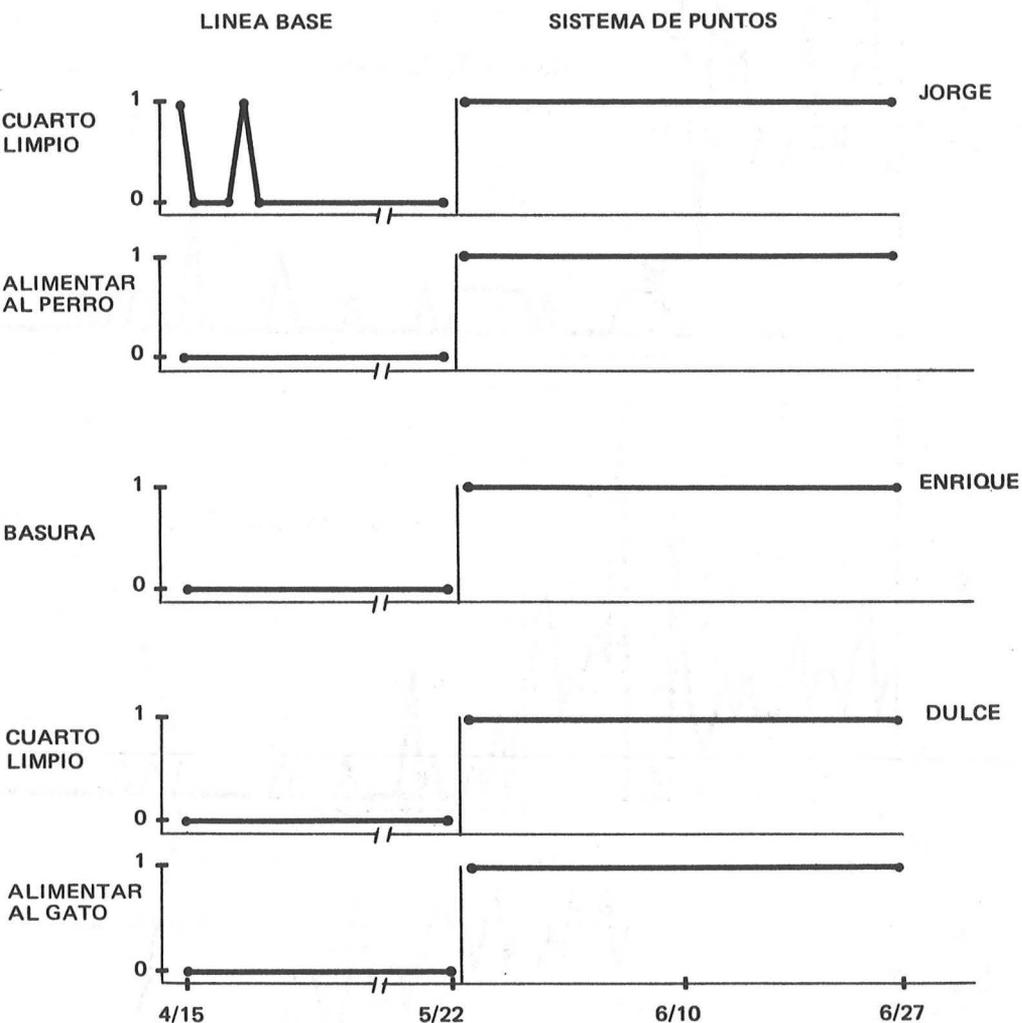


Figura 9. Efectos de la inclusión de un sistema de puntos para el mantenimiento de las conductas de tres niños (de 9, 5 y 8 años de edad), simultáneamente. (Tomado y adaptado de Christophersen, Arnold, Hill y Quilitch, *Journal of Applied Behavior Analysis*, vol. 5, núm. 4, verano de 1972, págs. 485-497. Copyright 1972 por la Society for the Experimental Analysis of Behavior, Inc.)

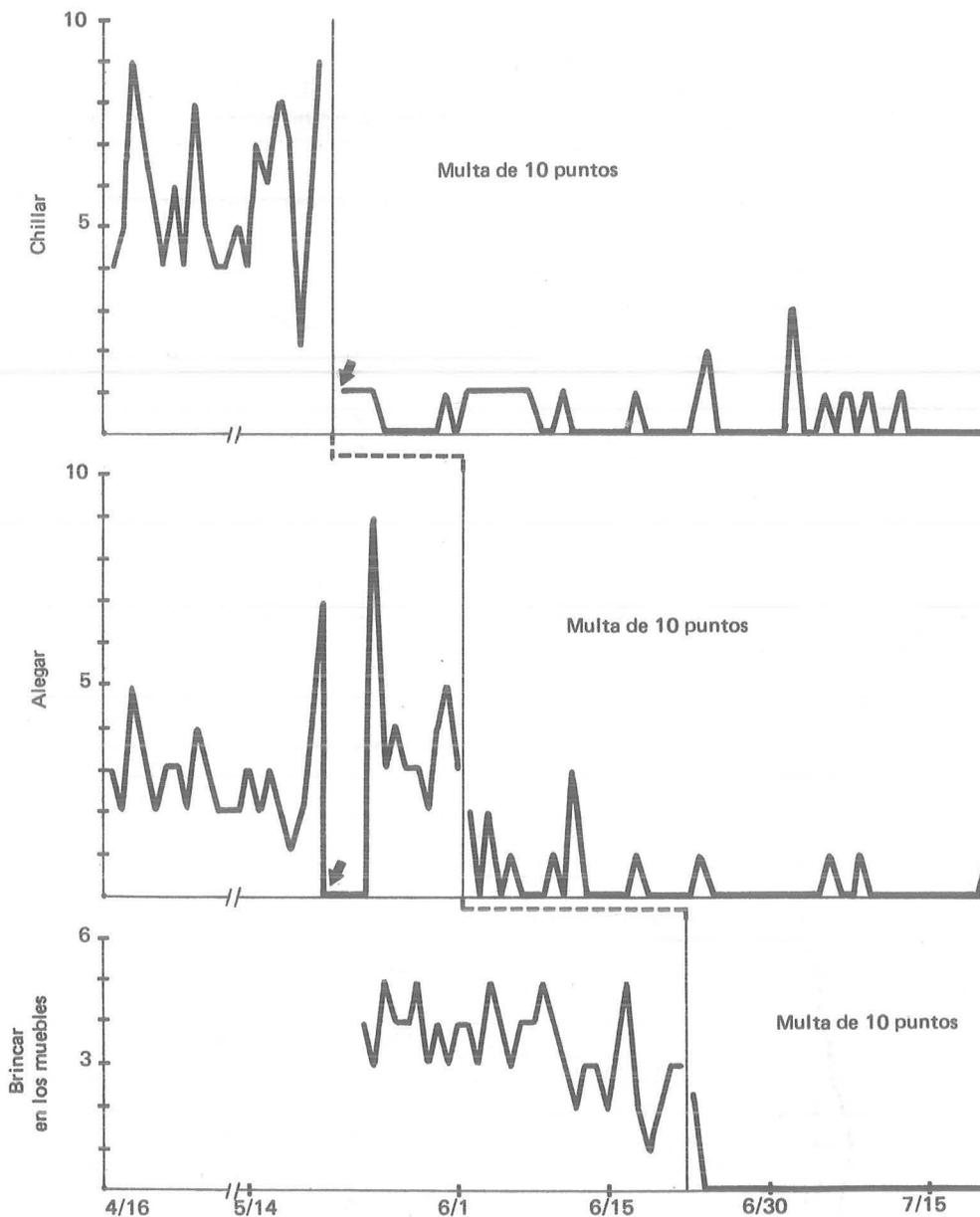


Fig. 10. Análisis de línea base múltiple de los efectos de la inclusión de multas de puntos sobre las conductas sociales de una niña de 8 años de edad. Las flechas indican cuándo se estableció el sistema de puntos para el mantenimiento de las conductas. (Tomado y adaptado de Christophersen, Arnold, Hill y Quilitch, *Journal of Applied Behavior Analysis*, vol. 5, núm. 4, invierno de 1972, págs. 485-497. Copyright 1972 por la Society for the Experimental Analysis of Behavior, Inc.)

Estos investigadores infieren que un proceso de aprendizaje vicario (adoptar un modelo) fue la causa principal de la diferencia observada. Desde el punto de vista del diseño experimental, se puede considerar que en realidad se tuvieron dos diseños A-B (uno para el grupo de niños con modelo y otro para el grupo de niños sin modelo) y que las VIs que se manipularon durante las fases B fueron diferentes.

El propio Bandura (1969) reporta un estudio en el cual se empleó el diseño A-B, elaborado por Philbrick y sus colaboradores. En dicho estudio se registró una línea base (con frecuencias bajas de respuestas correctas) de respuestas dadas en bloques de palabras antes de la verbalización de un principio útil, a fin de obtener ejecuciones correctas (A), y después de dicha verbalización (B). El incremento de una etapa a la otra fue sumamente marcado.

Hemmes y Eckerman (1972) estudiaron la interacción de dos diferentes programas de reforzamiento; después de someter a varias palomas a un programa múltiple de intervalo variable-intervalo variable (ambos con el mismo parámetro, por ejemplo, 3 minutos, 3 minutos), uno de los componentes (el correspondiente a la luz verde) fue cambiado a un programa *rda* (de reforzamiento diferencial alto), mientras el otro (luz roja) permaneció constante. El procedimiento utilizado obedece a un diseño A-B, en el cual la fase A corresponde al programa *múlt* IV IV y la fase B corresponde a la combinación *múlt* IV *rda*. Los resultados de este estudio se sintetizan claramente en la figura 12.

Como se puede ver, en los cuatro primeros animales el programa de *rda* tuvo el efecto de incrementar la tasa de respuesta del componente, que permaneció constante bajo el programa de IV. Hemmes y Eckerman no pudieron explicar este hallazgo, pero sí formularon dos explicaciones potenciales: *a*) es posible que haya ocurrido un cambio en la distribución de los intervalos interreforzamiento del componente variable simultáneo a la inclusión de *rda*, produciendo así una mayor tasa de respuesta, o *b*) el cambio observado en el componente IV simplemente representa una generalización de estímulo producida por *rda*.

Algunos otros estudios se han efectuado con este diseño por: Wolf y colaboradores (1964), Meichenbaum y colaboradores (1968), Whitlock (1966), Tooley y Pratt (1967), Straughan y colaboradores (1965), Solyon y Miller (1965), Rafi (1962), Quay y colaboradores (1966), Perline y Levinsky (1968), O'Leary y Becker (1967), McKenzie y colaboradores (1970), Milby (1970) y Millard (1966). Para una revisión exhaustiva, se recomienda la bibliografía de Breyer y Axelrod (1971).

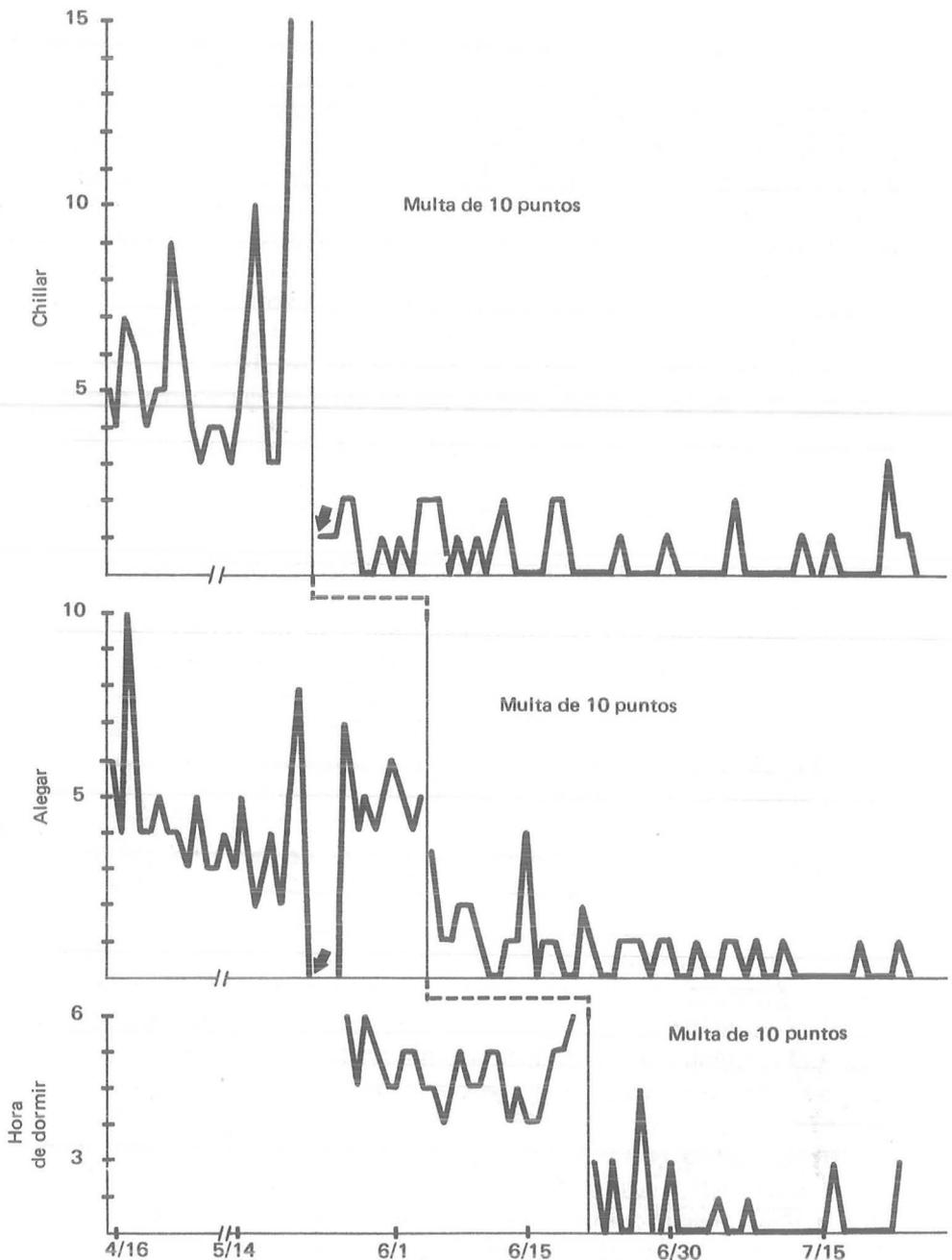


Figura 11. Análisis de línea base múltiple de los efectos de la inclusión de multas de puntos sobre las conductas sociales de un niño de 5 años de edad. Las flechas indican cuándo se estableció el sistema de puntos para el mantenimiento de las conductas. (Tomado y adaptado de Christophersen, Arnold, Hill y Quilitch, *Journal of Applied Behavior Analysis*, vol. 5, núm. 4, verano de 1972, págs. 485-497. Copyright 1972 por la Society for the Experimental Analysis of Behavior, Inc.)

dades experimentales son sometidas a las condiciones experimentales. Una medida tendiente a controlar esta variable consiste en someter a algunos organismos al orden A-B y a otros al orden B-A. Si el orden no es fundamental, entonces los datos en ambas fases *A* deben ser equivalentes entre sí, y los datos en las fases *B* también deben ser equivalentes entre sí. Otra forma de plantear este problema sería considerar que los efectos de la manipulación experimental ocurridos en la primera etapa son “arrastrados” hasta la segunda y que lo que verdaderamente se está apreciando en ésta es el efecto de la VI, más los efectos retardados de la primera.

En un experimento simple sobre condicionamiento clásico, Castro y Escabi (1973, datos inéditos) utilizaron un diseño balanceado, a fin de controlar el orden de presentación del estímulo incondicionado (un limón partido por la mitad) y el del estímulo condicionado (una tarjeta con la palabra “limón”). Se midió la salivación de dos sujetos ante la presencia de estos dos estímulos (después de medir su salivación normal). Los datos, en mililitros, fueron los siguientes:

$$\begin{array}{l} S_1) \quad T = 1.8 \quad L = 3.6 \\ S_2) \quad L = 4.6 \quad T = 2.6 \end{array}$$

En este caso se observa que los estímulos incondicionados produjeron casi el doble de salivación que los estímulos condicionados. Aunque se podría decir que cuando se presentó primero el estímulo incondicionado (orden B-A) esta relación no fue tan estrecha, la diferencia parece ser lo suficientemente grande como para indicar una mayor magnitud de la respuesta al estímulo incondicionado (lo cual está de acuerdo con lo observado en estudios más relevantes en este campo).

Hope (1970) señala que un diseño balanceado como el de este tipo, en el cual cada sujeto sirve como su propio control, es un diseño “deshonroso”. Cita como principio profiláctico para contrarrestar las características de “balanceo” y “propio control”, que no es lo mismo darle a un sujeto una prueba de inteligencia y luego dispararle un tiro... que primero dispararle y luego darle la prueba.

Ahora bien, pasando por alto lo absurdo de la “experimentación” que este autor trata de ilustrar, refirámonos con un poco más de detalle a esta cuestión. Suponiendo que las puntuaciones de CI fueran una VD razonable y que los tiros fuesen mortales, el sujeto medido antes de dispararle, obtendría un CI de, por ejemplo, 120; y el sujeto medido después de dispararle obtendría una puntuación de 0. El diseño permitiría apreciar el efecto (un tanto impresionante en este caso) que ejerce el orden de nuestras “operaciones” sobre el CI. Aunque el diseño no suprimiría los efectos del orden, sí serviría para apreciarlos.

6.3. DISEÑO DE APAREAMIENTO AYUNTADO

Apareamiento ayuntado. En este diseño se aparean dos organismos, estableciéndose alguna forma de control de uno de ellos sobre la conducta del otro. Un ejemplo de dicho procedimiento lo encontramos en un trabajo de Killen (1972). En una caja con dos *operandos*, este autor entrenó a una paloma a responder bajo un programa concurrente; cuando una de las claves era verde, los reforzamientos correspondían a un subprograma de IV (intervalo variable), y cuando era roja correspondían a otro subprograma de IV. Picar la segunda clave producía el cambio de color de la primera clave y, por tanto, el cambio de subprograma.

La segunda paloma estaba en una caja con un solo *operando* funcional y bajo un programa múltiple que constaba de dos subprogramas de IV equivalentes a los del programa concurrente. El color de la clave era exactamente igual al que tuviera la primera caja y cambiaba de acuerdo a ésta.

El programa múltiple estaba ayuntado al (dependía del) programa concurrente, ya que el estímulo discriminativo de la segunda cámara variaba de acuerdo a la forma en que lo hacía su contraparte de la primera caja. También la presentación de reforzamientos dependía de la conducta de la primera paloma.

Se observó, entre otras cosas, que las palomas en el programa múltiple ajustaron en cierta forma sus tasas de respuesta a las de las palomas bajo el programa concurrente. Se indicó que los colores del disco translúcido (que dependían de la primera paloma) ejercieron cierto control sobre la conducta de la paloma ayuntada (dependiente).

Se ha señalado (Church 1964, D'Amato, 1970) que la principal desventaja de este tipo de diseño consiste en que puede producir situaciones favorables o desfavorables para los sujetos ayuntados, ya que (en la mayoría de los casos) sus parámetros de respuesta serán diferentes de los de los sujetos experimentales, y su ejecución puede distorsionarse peligrosamente. Quizá habría que agregar que en un gran número de problemas esto es precisamente lo que se trata de producir cuando se emplea este diseño, ya que si se quisiera estudiar la conducta del organismo sin ayuntar, simplemente se le analizaría por separado.

De cualquier manera, ya se ha señalado (Kimmel y Terrant, en D'Amato, 1970) el desarrollo reciente de un diseño acoplado recíproco, a fin de suplir algunas de las deficiencias del apareamiento ayuntado.

II. CLASE TRADICIONAL

6.4. DISEÑO DE DOS GRUPOS ALEATORIOS

Este tipo de diseño implica, generalmente, un arreglo con dos condiciones: una condición o grupo control, y una condición o grupo experimental. A dicho arreglo lo representaremos con la connotación C-E. También es posible que haya dos condiciones experimentales en este diseño, sin incluir un grupo control. Dicho arreglo lo representaremos como E_1 - E_2 , indicando las dos condiciones experimentales que se obtienen de dar dos valores diferentes de la VI a dos grupos de sujetos. Este diseño también se ha denominado *bivalente* (Plutchik, 1968), por incluir dos valores de la VI.

Las características principales del diseño de dos grupos aleatorios son: *a*) se utiliza el procedimiento de aleatorización para asignar las unidades experimentales a las condiciones; *b*) los sujetos se asignan exclusivamente a una de dos condiciones experimentales; *c*) generalmente, sólo se mide una vez a cada sujeto por cada VD (cuando se mide más de una ocasión a cada sujeto, se considera que se trata de un diseño de mediciones repetidas); *d*) la información que se puede obtener es escasa, y *e*) dentro de los diseños de este tipo, es el que menos exigencias tiene, en términos de número de sujetos, equipo, ayudantes, etc.

El diseño de dos grupos aleatorios se usa frecuentemente en la investigación exploratoria. Sirve para indicar, en una forma gruesa, si la VI ejerce algún efecto en la VD; y si lo ejerce, cuál es su dirección (si tiende a haber una relación directa o si tiende a haber una relación inversa); así como su magnitud aparente. El diseño puede producir información útil, que puede aprovecharse para el planeamiento de experimentos más complejos. En ocasiones, la información que se obtiene puede ser no sólo muy escasa, sino incluso engañosa, como se verá posteriormente.

Rovner (1972) usó un diseño C-E en un experimento acerca del efecto de primacía en la memoria a corto plazo, en niños. Mandel y Bridger (1973) utilizaron un diseño de dos grupos aleatorios para investigar la transferencia del condicionamiento de una modalidad sensorial a otra. Otros estudios en los que se ha empleado alguna forma de este diseño se han llevado a cabo por: McAllister y colaboradores (1969), Brown y Foshee (1971), Clement y colaboradores (1970), Farina y colaboradores (1969), Foreyt y Kennedy (1971), Krapfl y Nawas (1969), Litvak (1969), Meyer y colaboradores (1970) y Cermak y colaboradores (1972).

El fundamento mismo del diseño es la equivalencia que, se supone, se logra mediante la aleatorización. Ya se ha expuesto lo suficiente acerca del razonamiento en que se basa este proceso. Se le considera como un medio, independiente del experimentador, de hacer equivalentes a los grupos antes de la experimentación. Hay dos desventajas principales al emplear este procedimiento, a saber: *a*) la pseudoaleatorización, y *b*) la posibilidad, poco probable, de inequivalencia. El primer problema surge cuando el procedimiento de asignación de los sujetos a los grupos no es verdaderamente al azar; por ejemplo, si se utiliza un procedimiento "casero" para implementar la aleatorización, se puede incurrir en toda clase de irregularidades. Cuando se cortan pequeños pedazos de papel con los nombres de los sujetos, la asignación de éstos a sus grupos no es realmente independiente de la persona que efectúa el sorteo. Lo mismo sucede si se ordena alfabéticamente a los sujetos. Se sabe que incluso algunas de las tablas de números aleatorios producidas por computadora son, realmente, números pseudoaleatorios. Se puede considerar, a pesar de esto, que las tablas producidas por procedimientos electromecánicos constituyen la forma más confiable de aleatorización.

El segundo problema no ha sido estudiado con suficiente cuidado en la literatura especializada. McGuigan (1972), por ejemplo, reduce este problema a una regla de muestreo: "emplear un número suficientemente grande de sujetos". Aunque ésta es una manera bastante segura de reducir la probabilidad de que ocurra un accidente de inequivalencia, ésta siempre existirá y podría invalidar un experimento. Dingman (1969) solía asignar a sus estudiantes la simulación de un "experimento" en una computadora. Los "valores de la VD" eran obtenidos de tablas de números aleatorios que eran divididos arbitrariamente en "condiciones experimentales". A continuación se analizaban estadísticamente, y un número cercano al 3% de estos "experimentos" producía diferencias "significativas". El método es sencillo: no importa cuán pequeña sea, siempre existe la probabilidad de producir grupos desiguales cuando se utiliza el procedimiento de aleatorización. A fin de evitar esta posibilidad, se sugiere replicar el experimento un gran número de veces.

Quizá la desventaja principal del diseño de dos grupos aleatorios es su falta de sensibilidad. Supongamos que un experimento con este diseño indica que hay una relación directa entre la VI y la VD, como se señala en la figura 13. Sin embargo (y seguimos suponiendo), si se hubiese ejecutado un experimento con más de dos grupos, se habría encontrado que dicha relación directa solamente era válida para el rango más bajo de la VI (como lo indica la línea continua de la misma figura). Este es un ejemplo de insensibilidad del diseño. Otra posibi-

lidad se ilustra en la figura 14; en este caso, el experimentador (hipotético) decidió estudiar dos valores extremos de la VI. De acuerdo a la evidencia proporcionada por este diseño, prácticamente no habría ninguna diferencia entre ambos tratamientos y, por ende, ningún efecto de la VI sobre la VD. Sin embargo, si el investigador hubiese podido estudiar algunos de los valores intermedios en el rango de la

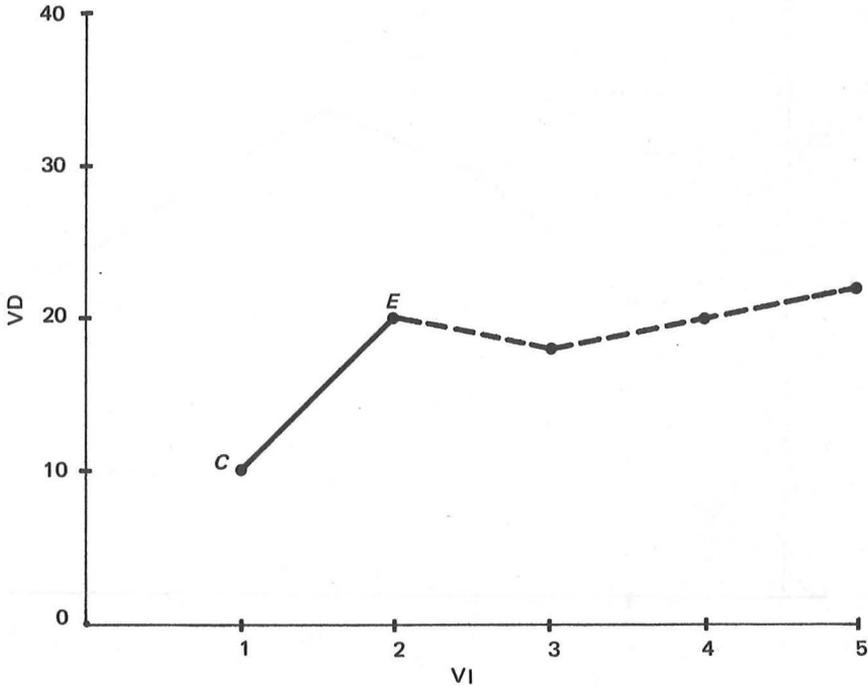


Figura 13. Datos ficticios que muestran una relación lineal directa entre la VI y la VD en el rango limitado a dos condiciones experimentales (línea continua) correspondientes a un *diseño de dos grupos aleatorios*. Esta relación no se sostiene en el resto del rango de la VI (línea punteada).

VI, habría encontrado que sí existe un efecto de la VI (línea punteada de la figura 14) y una relación funcional obvia entre ambas variables.

Un tercer caso de insensibilidad del diseño se muestra en la figura 15. En este caso, el diseño sería E_1-E_2 y, tal como la línea continua de la gráfica sugiere, se podría inferir erróneamente que mientras mayor sea el número de días de privación, menor será la tasa de respuestas. Aunque esto es válido para el extremo superior de la VI, exactamente la relación opuesta se encontraría en el extremo inferior (línea punteada), pero el diseño no contribuiría a apreciar este hecho. En todos

estos casos, el investigador puede caer en el error de sobregeneralizar, como acertadamente señaló Plutchik (1968).

Otras desventajas del diseño son su dependencia de muestras grandes y la incompleta información que proporciona. Por otra parte, las ventajas principales son su simpleza y economía.

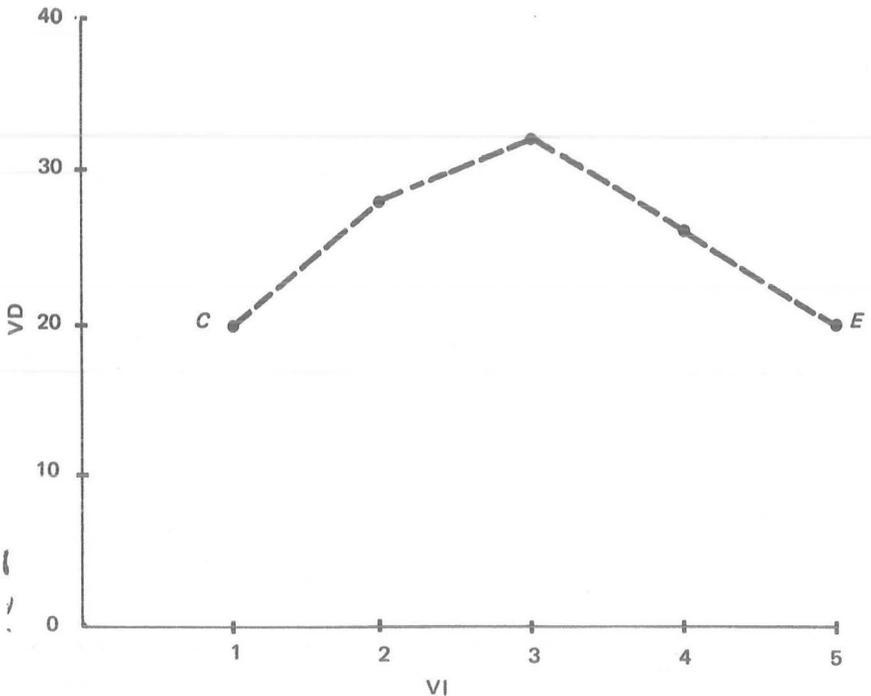


Figura 14. Datos ficticios que muestran una ausencia aparente del efecto de la VI sobre la VD. Nótese que no hay diferencia entre los valores de la VD en los puntos C y E.

6.5. DISEÑO DE APAREAMIENTO DIRECTO

Parece ser que, originalmente, el procedimiento de aparear las unidades experimentales era *a posteriori*. Fisher (1966) lo propuso a principios de este siglo en estudios de agricultura como una forma de llevar a cabo análisis de datos.* Se tomaba a la planta más alta del grupo experimental y se le apareaba con la más alta del grupo control; se tomaba a la segunda planta más alta del grupo experimental y se

* Fisher experimentó en agricultura para el AVAR. [N. DEL R.]

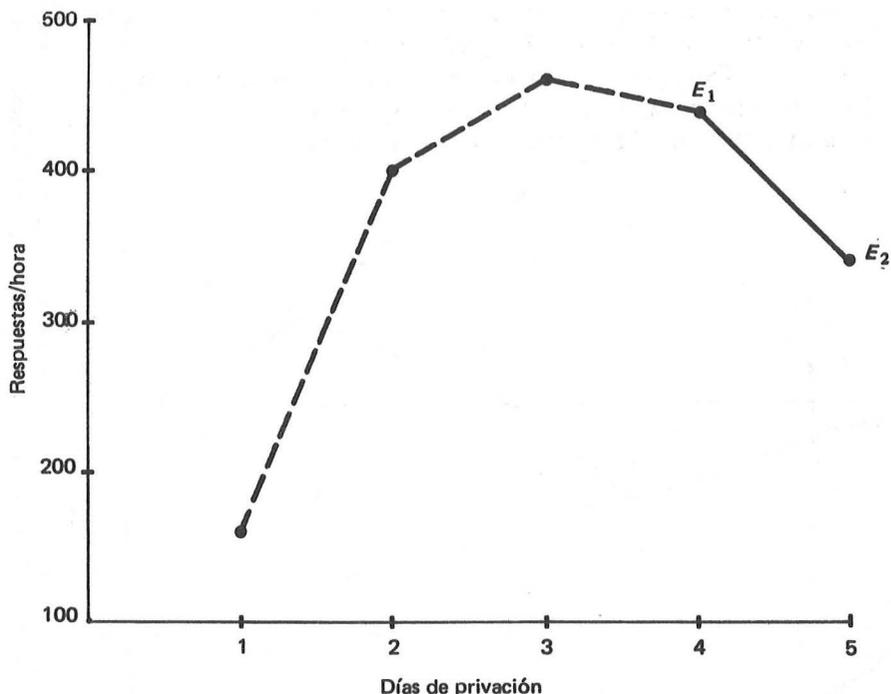


Figura 15. Datos ficticios que muestran una aparente relación inversa entre la VI (días de privación) y la VD (tasa de respuesta), sugerida por el uso de un *diseño de dos grupos aleatorios* (línea continua).

le apareaba con la segunda más alta del grupo control, y así sucesivamente, hasta comparar todos los pares de plantas disponibles. En esta forma, se podía observar con cierta facilidad si el grupo experimental daba valores (en la VD) sistemáticamente mayores que los del grupo control. Como ésta era una técnica rudimentaria de análisis de datos, posteriormente fue reemplazada por otras más eficaces.

Previamente analizamos los diseños preexperimentales de apareamiento indirecto; ahora mencionaremos algunos diseños verdaderamente experimentales de apareamiento directo. En éstos, el apareamiento es un procedimiento de diseño *previo a la inclusión de la VI*, y no un procedimiento de análisis de datos, como el de Fisher. Plutchik (1968) hizo una detallada exposición de estos diseños, y aquí adoptaremos parte de su clasificación:

a) *Apareamiento por características observables.* En este diseño se mide alguna característica objetiva de los sujetos, como estatura, peso, edad o sexo; por ejemplo, si la VA fuese estatura, se asignaría

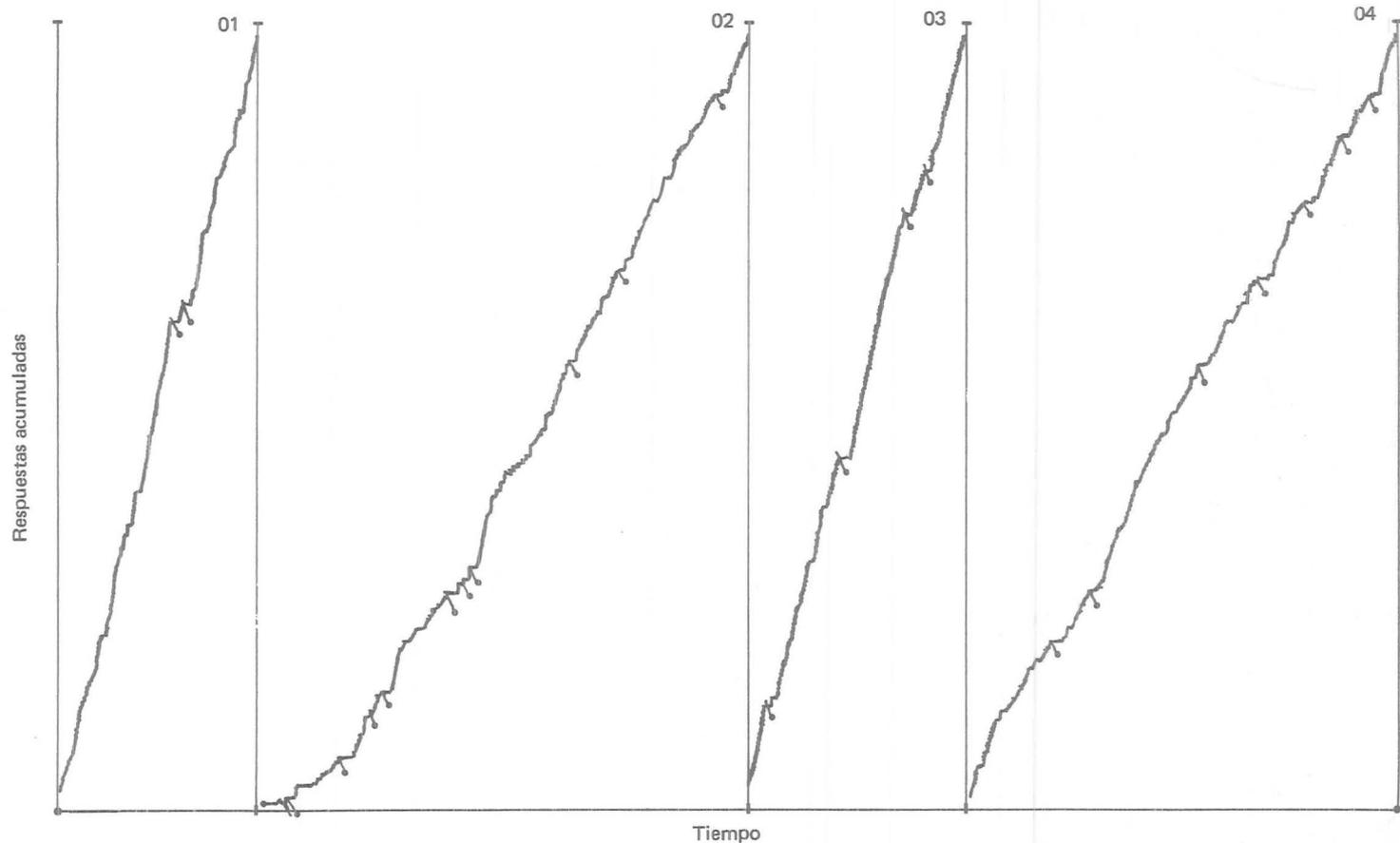


Figura 16. Datos ficticios que muestran la ejecución de cuatro organismos (O) diferentes, la cual sirve de base para aparearlos. Teniendo en cuenta la pendiente de las líneas y el tiempo invertido el organismo O_3 se debería aparearse con el O_1 , y el O_4 con el O_2 .

en una forma aleatoria a uno de los dos sujetos más altos al grupo experimental, y el otro se asignaría al grupo control. Después se tomarían los dos siguientes y se repetiría el proceso. De esta manera, hay bastante seguridad de que los grupos sean equivalentes en esta variable, antes de la aplicación de la VI. Esto es, por supuesto, teniendo en cuenta que la variable de estatura sea relevante al comportamiento de la VD.

b) *Apareamiento por ejecución.* Este diseño representa el caso tan temido que originó el diseño de apareamiento indirecto-correlacional. Cuando aún no se empleaba la tecnología de la línea base, se creía, fundadamente, que si se medía al sujeto (generalmente a través de su ejecución) en la VD a ser afectada por la VI, los resultados podrían confundir el efecto de ésta con el de la previa medición (que sería reactiva en este caso). Con el advenimiento de la línea base, este problema pasó a ser trivial. Si se hace que varios organismos sigan determinada conducta bajo el mismo programa de reforzamiento o castigo y se obtiene una verdadera línea base, entonces se puede aparear a aquellos que presentan líneas base similares, se aplica la VI y se evalúan claramente sus efectos por medio de dos comparaciones: la de la nueva ejecución contra la línea base anterior, y la que se obtiene al contrastarla con la ejecución del organismo testigo. (La figura 16 muestra, con datos ficticios, una forma de apareamiento por ejecución.)

7

diseños univariantes multicondicionales

EN ESTA CATEGORÍA se incluyen aquellos diseños que presentan más de dos condiciones experimentales. Además, a pesar de que el número de condiciones pudiera llegar a ser relativamente considerable, en estos diseños solamente se explora una variable independiente. Esto significa que algunos diseños de esta clase permiten investigar un rango relativamente grande de la VI; incluso, en ocasiones se puede cubrir el rango de la VI exhaustivamente. Otros tipos de diseños que aquí estudiaremos sólo se refieren a unos cuantos valores de la VI, pero sirven al propósito de demostrar inequívocamente el efecto que ejerce la VI sobre la VD y, asimismo, un alto grado de control sobre la conducta del organismo estudiado.

I. CLASE CONDUCTUAL

7.1. DISEÑO REVERSIBLE A-B-A

Quizá la mejor descripción (aunque solamente implícita) de este diseño es la que dan Ferster y Skinner (1957): "...establecer una

línea base, incluir una variable (independiente)* y retornar de nuevo a la línea base, de tal manera que el organismo sirva como su propio control" (pág. 38).

A esta operación del experimentador de volver a la línea base original pudiera corresponder un cambio en el mismo sentido de parte de la conducta del organismo estudiado. Cuando esto sucede, se dice que la conducta es reversible o que se ha recuperado la conducta original; de ahí el nombre de diseño reversible. Sin embargo, habrá ocasiones en que el experimentador suprime la VI, volviendo a aplicar las condiciones que produjeron la línea base original (fase A), y la conducta podría no regresar a su estado original. En este caso, se dice que la conducta es irreversible o parcialmente irreversible.

Hasta principios de la década pasada varios de los autores más sobresalientes en el campo del AEC (Ferster y Skinner, 1957; Sidman, 1960) afirmaban que era poco probable encontrar estados conductuales completamente irreversibles; sin embargo, el propio Sidman reporta por lo menos un caso de irreversibilidad parcial (un estudio de Boren, págs. 242 y 253). Además, se podría añadir que en algunos estudios fisiológicos se producen verdaderamente estados irreversibles, por ejemplo, cuando se producen lesiones severas en partes del cerebro o en el sistema nervioso (como cerebeloectomías, lobotomías, o separación de los hemisferios cerebrales). Comoquiera que sea, Sidman advierte que el experimentador debe estar preparado para tratar con una verdadera irreversibilidad, si alguna vez la encuentra, y que ésta no puede ser nulificada por ningún diseño experimental, ni mediante control experimental o estadístico.

El prototipo de este diseño tipificó, según Campbell y Stanley (1966), gran parte de la investigación científica del siglo XIX en el campo de la física y la biología. Estos autores proporcionan la siguiente ilustración: se toma un gran número de mediciones (fase A) del peso de una barra de hierro; luego, se sumerge ésta en ácido nítrico (fase B) y finalmente se le remueve, efectuando otra serie de mediciones (fase A). La pérdida de peso sufrida por la barra de hierro se atribuye, inequívocamente, al ácido nítrico. Hacen notar, además, que se consideraba innecesario tener barras testigo, aunque éstas pudieran estar presentes en el laboratorio.

Un experimento análogo podría ser efectuado por un psicólogo (y, de hecho, todavía se realiza esta clase de experimentos como demostraciones a estudiantes noveles), utilizando el mismo diseño; por ejemplo, podría estudiar la frecuencia con que una lagartija se trepa

* Adición del autor

a una vara cubierta con un cable eléctrico sin corriente (fase A₁). Entonces, se electrifica el cable y se registra la frecuencia con que la lagartija trepa a la vara (fase B). Se retira al animal de la situación experimental por unos días y, finalmente, se le regresa al escenario original no electrificado (fase A₂). Se registra la frecuencia con que el animal vuelve a trepar la vara y, si ésta es tan baja como la observada durante la fase B, entonces se infiere que ha ocurrido un proceso de aprendizaje.

Hoy día, en el AEC, es común el uso de este diseño, y se ha empleado desde hace mucho tiempo. Skinner (1938), por ejemplo, reportó varios experimentos en los que se obtuvieron diferentes formas de recuperación, utilizando en todas ellas el diseño A-B-A (Underwood ya lo citaba en 1949). Logan y Wagner (1965) citaron un experimento de Holz y Azrin en el que, después de pasar por diferentes programas de reforzamiento, sometieron a un grupo de palomas a una fase de castigo (A), en la cual la tasa de respuesta fue considerablemente baja; después se les suspendía el castigo durante 10 minutos (B), observándose un marcado aumento en la tasa de respuesta; finalmente, se reimplantaba el castigo (A), disminuyendo de nuevo la tasa de respuesta.

Johnston y Johnston (1972) utilizaron este diseño en combinación con un registro de línea base múltiple, en una investigación cuyo objetivo era modificar la articulación de los sonidos consonantes en niños pequeños. Durante la línea base (A) de su primer experimento, se observó que el nivel de sonidos correctos era sumamente bajo, en comparación con el nivel de los sonidos incorrectos. Durante la fase manipulativa de reforzamiento (B), se les indicó a los niños que se estaban registrando los sonidos que emitían y se les hizo ver que "ganarían" diferentes símbolos por los sonidos de consonantes correctamente emitidos. Estos símbolos eran acumulados en gráficas individualizadas y, una vez que rebasaban cierto nivel, podían escoger con qué jugar. Estos investigadores informan haber logrado invertir los niveles de sonidos correctos e incorrectos por medio de dicho procedimiento. Finalmente, se reincluyeron las condiciones prevalecientes durante la línea base (A). Entonces se produjeron incrementos (desde moderados hasta grandes) en las tasas de respuestas incorrectas y decrementos similares en el número de respuestas correctas. En otro experimento se logró mantener el cambio obtenido en la fase B, mediante la generalización de estímulos que se produce al hacer discriminativo a un niño estímulo, con el fin de que otro niño produzca una articulación correcta. La figura 17 presenta los datos obtenidos con el diseño A-B-A.

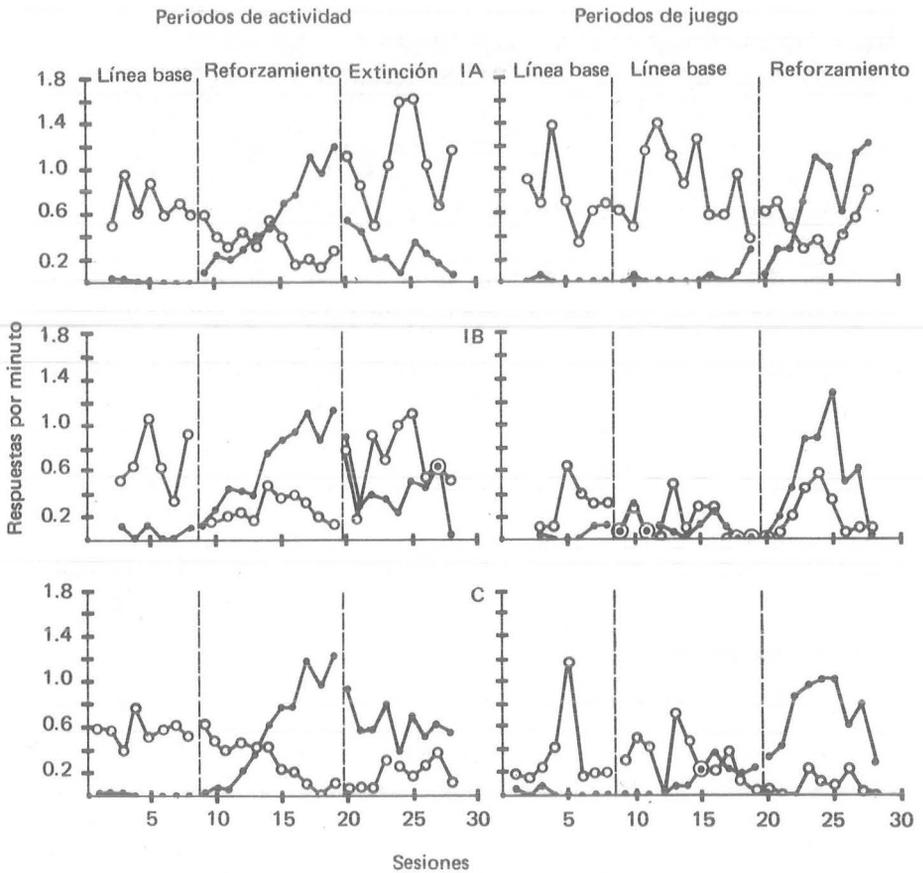


Figura 17. Número de sonidos emitidos, correctos e incorrectos, por minuto para todos los sujetos en el experimento 1. Cada punto representa las tasas totales correctas para todos los periodos de actividad o juego en cada sesión diaria. Los puntos sólidos indican los sonidos correctos, y los puntos en blanco representan los sonidos incorrectos. IA, IB e IC son tres diferentes sujetos. (Tomado y adaptado de Johnston y Johnston, *Journal of Applied Behavior Analysis*, vol. 5, núm. 3, otoño de 1972, págs. 233-246. Copyright por la Society for the Experimental Analysis of Behavior, Inc.)

Como parte de un procedimiento más complicado, Weisman y Ramsden (1973) realizaron un experimento en el que utilizaron el diseño A-B-A. Durante la fase A (después de un entrenamiento bajo un programa de razón fija), las palomas fueron sometidas a un programa múltiple de *intervalo variable 1 min, intervalo variable 1 min*. Aunque un estímulo (E_1) estaba correlacionado con uno de los componentes (IV 1 min) y el otro (E_2) con el segundo componente (IV 1 min),

no se les reforzó diferencialmente en esta fase. Se observaron niveles de respuesta similares para ambos estímulos.

Durante la fase B, se cambió a un programa múltiple de *intervalo variable de 1 min, tiempo variable 1 min* (IV 1 min, TV 1 min) con el fin de establecer una discriminación entre los dos estímulos.

Cuatro palomas tuvieron al E_1 (una línea con 0° de inclinación) relacionado con el primer componente (IV 1 min) y al E_2 relacionado con el segundo componente del programa múltiple (TV 1 min). Otras cuatro palomas tuvieron el orden opuesto.

Durante esta etapa se observó un decremento en las tasas de respuesta controladas por E_2 (indiferentemente de su forma física).

Finalmente, se reincluyeron las condiciones de la fase A (programa múltiple IV 1 min, IV 1 min) y se recuperaron las tasas de respuesta originales (véase la figura 18).

Otro ejemplo del uso del diseño A-B-A en el campo de la investigación conductual básica es el trabajo de Downey y Harrison (1972). Estos investigadores condicionaron a unos monos a responder a un operando adyacente a una bocina que emitía un sonido; por cada respuesta emitida en el operando adyacente recibían reforzamiento (fase A). Se observó un alto porcentaje de respuestas "correctas" durante esta etapa.

Durante la etapa B, los monos fueron sometidos a condiciones "no adyacentes", es decir, a condiciones en las cuales las bocinas quedaban alejadas a los operandos para responder. En la figura 19 se puede observar cómo el porcentaje de respuestas correctas disminuyó notablemente en esta fase.

Finalmente, se les regresó a las condiciones adyacentes originales (etapa A) y se observó una recuperación del nivel de la línea base a un punto comparable al de la primera fase (A).

Phillips y sus colaboradores (1973) utilizaron este diseño para comparar tres diferentes grupos de muchachos con conductas delictivas. Algunos otros investigadores que han reportado trabajos en los que han utilizado el diseño A-B-A son: Peterson y Whitehurst (1971), Craig y Holland (1970), Barton (1970), Azrin y colaboradores (1968), Hake y Campbell (1972), Wincze (1972), Koegel y Covert (1972) y Waite y Osborne (1972).

Los ejemplos aquí presentados podrían servir para juzgar la gran variedad de situaciones en las cuales se ha empleado este diseño reversible. Es importante señalar que el nombre de "diseño reversible ABA" no es de uso general. En ocasiones sólo se le refiere como diseño reversible; a veces se habla de "línea base-tratamiento-línea base" o de "línea base-experimentación-línea base"; en otras ocasiones se emplean

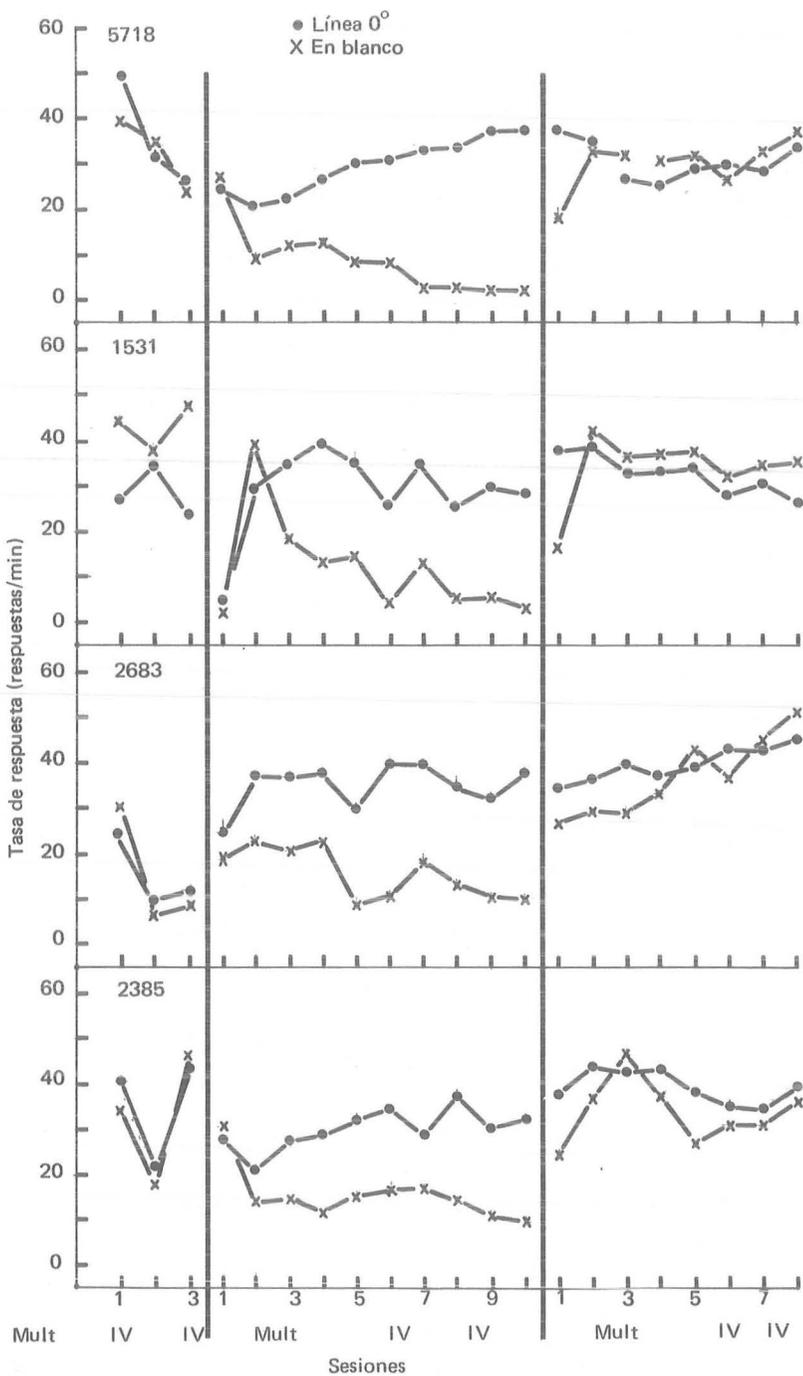


Figura 18. Tasas de respuesta a la línea de 0° y a la clave verde en blanco (apagada) durante el reforzamiento diferencial y no diferencial en el experimento 1. La clave verde en blanco estuvo correlacionada con el reforzamiento IV. (Tomado de Weisman y Ramsden, *Journal of the Experimental Analysis of Behavior*, vol. 19, núm. 1, 1973, págs. 55-64. Copyright 1973 por la Society for the Experimental Analysis of Behavior, Inc.)

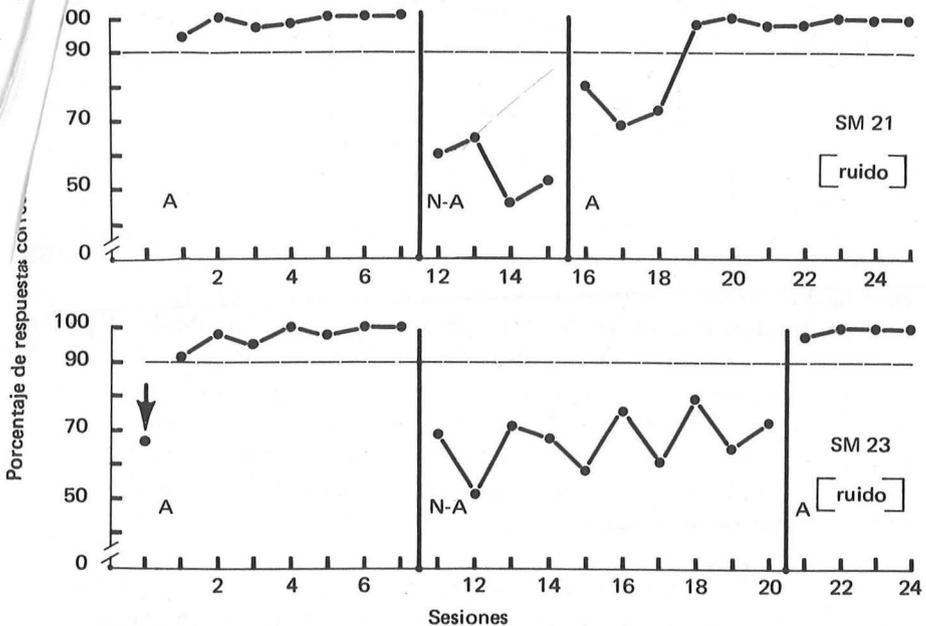


Figura 19. Adquisición del control de estímulos sobre las respuestas de unos monos, obtenida con las bocinas en la posición adyacente (A), y efectos de un cambio posadquisición a la posición no adyacente (N-A) seguidos por un regreso a la posición adyacente (A). Se usó reforzamiento no diferencial durante las tres fases del diseño. Los estímulos fueron ruido o tronidos de 10-khz (indicados para los animales individuales entre paréntesis). El punto marcado con una flecha (en SM 23) representa la conducta antes de la inclusión de los estímulos de sonidos. (Tomado de Downey y Harrison, *Journal of the Experimental Analysis of Behavior*, vol. 18, núm. 3, 1972, págs. 453-463. Copyright por la Society for the Experimental Analysis of Behavior, Inc.)

los números romanos I y II para denominar a las etapas A y B; y en otras más, ni siquiera se menciona el diseño utilizado.

Consideremos ahora las implicaciones que tiene el uso de este diseño tanto en la investigación básica como en la aplicada.

En el caso de la investigación básica, el principal objetivo de hacer la reversión de condiciones es apreciar si efectivamente el cambio observado en la VD con la inclusión de la VI se debe a esta última, y no a alguna otra variable que pudiese haber operado contingentemente a la VI. Un segundo objetivo sería evaluar los efectos residuales retardados que ejerce la VI sobre la VD.

En el caso de la investigación aplicada, las implicaciones son diferentes. Supóngase que (usando el diseño A-B) un investigador ha

logrado eliminar una conducta considerada como "indeseable", y tratando una conducta "deseable" en su lugar. Ahora está considerando la posibilidad de hacer una reversión de las condiciones y extendiendo el diseño a uno A-B-A. Por un lado, queda su curiosidad de averiguar si realmente su tratamiento experimental fue el responsable de los cambios obtenidos; por otro, está la posibilidad de producir conductas emocionales "negativas" que podrían resultar como consecuencia de la reversión (por ejemplo, debido a la suspensión del reforzamiento positivo) y el riesgo de que dichas conductas se generalicen al experimentador en sí (y no al procedimiento de la reversión). Ésta sería una posibilidad de perder "el terreno ganado", no sólo desde el punto de vista del experimentador, sino desde el punto de vista del sujeto afectado, que, en última instancia, es el más importante en este tipo de investigación. Parece obvio que en una situación de este tipo, es preferible utilizar el diseño A-B y dejar que los investigadores en el laboratorio satisfagan el rigor científico deseable. Ribes (1972), al referirse al uso de técnicas de control, que incluyen a la reversión, coincide en este punto.

Sin embargo, hay situaciones que podrían ser la contraparte del problema analizado; por ejemplo, en una investigación se podría usar el diseño A-B-A haciendo énfasis en la reversión. Supóngase que en una primera etapa (A), se le refuerza al sujeto por la emisión de una conducta preferida por él y que, además, es considerada como socialmente deseable. Se observa una alta tasa de respuesta. Durante la etapa B se refuerza al sujeto por la emisión de una conducta incompatible no preferida, y se observa una repentina disminución de la primera conducta (fase B). Aquí, obviamente, sería adecuado reincluir las condiciones originales de la fase A, ya que, además de comprobar la efectividad de la VI, se estaría modificando una conducta en una dirección socialmente aceptable.

7.2. DISEÑO REVERSIBLE A-B-A-B

Este tipo de diseño es considerado por Gentile y sus colaboradores (1972) como una forma específica de la replicación directa intrasujeto (estudiada extensamente por Sidman, 1960; véase sección 2.7). Aunque este investigador describe al diseño reversible A-B-A-B (o diseño de replicación intrasujeto) en términos de la aplicación de alguna forma de reforzamiento durante las etapas manipulativas (A), dicha descripción se puede generalizar a cualquier procedimiento en el que durante las etapas A se registre una línea base *sin la presentación de*

la VI, y durante las etapas B se registre una línea base *contingente a la presentación de la VI*.

Se dice que este diseño constituye una replicación intrasujeto porque la segunda secuencia A-B (A_2-B_2), es una repetición o réplica de la primera secuencia A-B (A_1-B_1) efectuada en el mismo sujeto. De nuevo, aunque Gentile y sus colaboradores se refieren a un incremento en la frecuencia de las respuestas como consecuencia del reforzamiento en las fases manipulativas B, se puede hablar, en general, de un *cambio en la probabilidad de la respuesta o de parámetros temporales asociados a ésta*, en relación al nivel operante registrado durante las etapas de comparación A.

Un ejemplo claro del uso de este diseño es el trabajo de Hasazi y Hasazi (1972). Estos investigadores modificaron la conducta de escribir ciertos dígitos en forma invertida (por ejemplo, 41 en lugar de 14). Tales dígitos eran el resultado de sumas que hacía un niño de 8 años de edad. Durante la primera fase de línea base (A_1), la maestra calificaba con una X las sumas con dígitos invertidos como incorrectos efectuadas por el niño (las sumas correctas se calificaban con una C), y le proporcionaba a éste una etapa de "ayuda extra". Dicha ayuda consistía en la aplicación de algunos procedimientos de la enseñanza tradicional. Durante esta línea base se observó un gran número de inversiones; el niño efectuaba casi tantas inversiones como sumas tenía que hacer. A continuación se llevó a cabo el primer periodo experimental (B_1). Durante este periodo, todas las sumas se le calificaron como correctas (aunque estuvieran invertidas); las sumas que no estaban invertidas eran reforzadas inmediatamente por la maestra con una sonrisa, una palmadita en la espalda o afirmaciones verbales como: "esta suma está muy bien hecha". Después de unas cuantas sesiones experimentales se observó una marcada disminución en el número de inversiones. A continuación se volvieron a efectuar estas dos etapas, repitiéndose las operaciones recién descritas (fases A_2-B_2). Como se puede observar en la figura 20, la conducta del niño obedeció a la forma que había presentado previamente durante las etapas A_1-B_1 .

Otro ejemplo, en el campo de la investigación básica, lo constituye el trabajo de Leander (1973). Este investigador realizó una serie de experimentos, a fin de conocer los efectos que ejerce la privación de comida en las respuestas de evitación bajo el paradigma de posposición de choque, de Sidman, y bajo el paradigma de reducción de la frecuencia de choque, de Hermstein y Hineline.

En el experimento relevante a esta sección, Leander midió tres VD (líneas base): peso del animal, tasa de choques y tasa de respuesta. Los sujetos fueron dos ratas. Durante la fase A del experimento, se

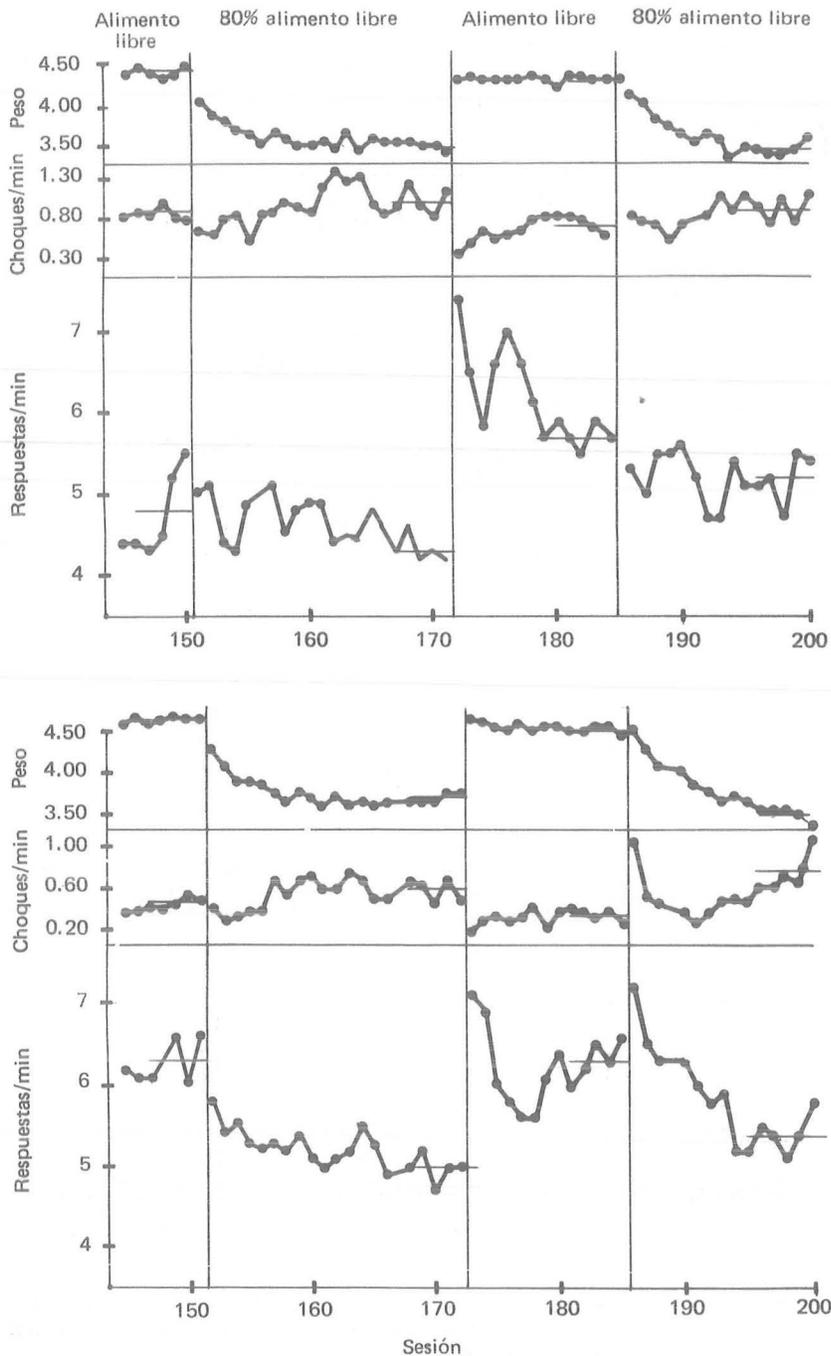


Figura 21. Tasas de respuesta, tasas de choque y peso corporal en gramos para las ratas CR 3 y CR 4 durante cada sesión de la fase 2. Las líneas verticales indican cuándo se iniciaron los cambios en las condiciones alimenticias. Las líneas horizontales muestran los valores medios de las tres variables dependientes para las últimas cinco sesiones. Diseño A-B-A-B. (Tomado de Leander, *Journal of the Experimental Analysis of Behavior*, vol. 19, núm. 1, 1973, págs. 17-24. Copyright 1973 por la Society for the Experimental Analysis of Behavior, Inc.)

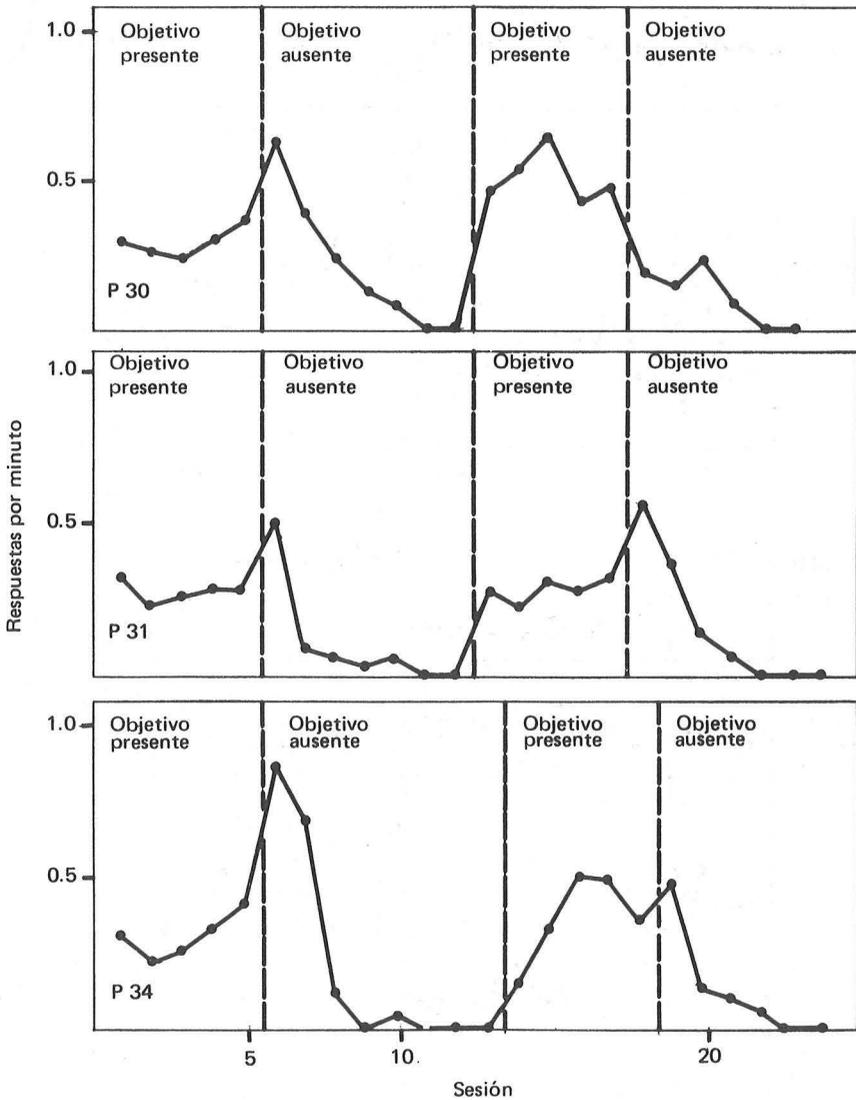


Figura 22. Tasa de respuesta en la clave objetivo en presencia y ausencia del objetivo. Diseño A-B-A-B. (Tomado de Cherek y colaboradores, *Journal of the Experimental Analysis of Behavior*, vol. 19, núm. 1, 1973, págs. 113-123. Copyright 1973 por la Society for the Experimental Analysis of Behavior, Inc.)

destruivas mediante el control del reforzamiento por los compaÑeros de clase, y el de Hall y colaboradores (1968) sobre los efectos que ejerce la atenci3n de los maestros en la conducta de estudio.

Hall (1971) hizo una descripci3n del diseÑo, breve y muy acertada. Aqu3 destacamos cuatro observaciones que efectu3 respecto a la forma de graficar los datos que se han adoptado com3nmente en la investigaci3n aplicada: *a*) las fases experimentales se dividen con l3neas verticales; *b*) los puntos entre condiciones adyacentes no se unen; *c*) los puntos de datos de alguna fase de seguimiento no se unen, y *d*) las condiciones experimentales se encabezan tan descriptivamente como sea posible. Ahora pasemos a una evaluaci3n del diseÑo.

Aun cuando deb3ramos conceder que una replicaci3n directa intra-sujeto (y, por ende, el diseÑo A-B-A-B) constituye una de las formas m3s contundentes de mostrar el efecto de cierta VI sobre la conducta de un organismo (de acuerdo con Skinner, Sidman, Bandura y muchos otros), esta aproximaci3n no deja de tener un n3mero de limitaciones desde el punto de vista del diseÑo experimental. Bandura (1969) hizo una de las cr3ticas m3s completas al respecto. De acuerdo con este autor, las principales limitaciones del diseÑo son: *a*) no puede emplearse cuando se produce un fen3meno conductual relativa o totalmente irreversible; *b*) no hay criterios cuantitativos (diferentes de los estad3sticos) ampliamente reconocidos por medio de los cuales se pueda evaluar incuestionablemente la efectividad de un cambio, sobre todo en los casos dudosos; *c*) el diseÑo reversible, en general, no permite, estrictamente hablando, la evaluaci3n de diferentes tratamientos (VI) aplicados de manera serial al mismo sujeto o grupo de sujetos, ya que no es posible separar los efectos de un segundo tratamiento de los efectos a largo plazo de un primer tratamiento, los de un tercero de los de un segundo, y as3 sucesivamente (v3ase el trabajo de Namboodiri en el cap3tulo 9, y tambi3n Campbell y Stanley, 1966); *d*) probablemente hay una diferencia fundamental entre la implantaci3n de una conducta por primera vez y su reimplantaci3n en ocasiones posteriores. Sidman (1960) desarroll3 un argumento bajo el mismo tenor, en el cual demuestra qu3 tan dif3cil puede ser lograr una verdadera reproducci3n de una l3nea base original, y *e*) este diseÑo, como otros reversibles, no evita la posibilidad de que una VE concomitante con la VI pudiera compartir con 3sta la "responsabilidad" por los cambios ocasionados.

Kazdin y Bootzin (1972) evaluaron, entre otros, el diseÑo A-B-A-B en relaci3n a un experimento mediante econom3a de fichas. Estos investigadores seÑalan que el diseÑo puede variarse, una vez que se ha demostrado una relaci3n funcional (por ejemplo, para apreciar la generalizaci3n de est3mulos o la resistencia a la extinci3n). Asimismo,

consideran que una de las principales ventajas de este diseño es la de eliminar al azar o a la maduración como fuentes de invalidez.

Al considerar los problemas que el diseño produce, concuerdan con Bandura en que éste es inaplicable cuando la conducta es irreversible o no transitoria. Específicamente, indican que la economía de fichas no siempre produce conductas reversibles. Otra limitación que señalan Kazdin y Bootzin es su inaplicabilidad cuando se prueba la resistencia a la extinción.

Una segunda categoría de problemas que encuentran estos investigadores consiste en que los efectos que producen las fichas (VI) se pueden confundir con cambios ambientales concomitantes con la presentación o retiro de las fichas. Como ejemplos específicos de dichos cambios están la conducta del personal o las instrucciones, como cuando se les dice a ciertos pacientes que (durante una reversión) "tendrán vacaciones pagadas", y se observa entonces una marcada disminución en el número de respuestas.

Finalmente, Kazdin y Bootzin señalan que los cambios obtenidos mediante este diseño en un escenario pudieran no generalizarse a distintos escenarios. Incluso, indican la posibilidad de que las reversiones pudiesen servir como un requisito para obtener recuperaciones exaltadas de la conducta. A pesar de todas estas limitaciones, Kazdin y Bootzin aceptan que el diseño "proporciona el elemento evaluativo más práctico para valorar programas en curso" (pág. 365).

Comoquiera que sea, no debe olvidarse que éste es uno de los diseños que ha dado lugar a valiosa información experimental y que sus méritos sobrepasan a sus limitaciones. Debemos considerar que el investigador que emplea este diseño, junto con un riguroso control experimental, puede llegar a demostrar tan claramente la forma en que una variable controla a otra, como una persona puede demostrar de qué modo un contacto eléctrico enciende o apaga una luz. Por otra parte, una replicación fallida puede controlar el comportamiento experimental de un investigador (sus procedimientos) tanto como una fuerte crítica de un colega. Además, ¿cómo saber si una conducta es reversible, o un hallazgo replicable, si no es a través de diseños de este tipo?

7.3. DISEÑO REVERSIBLE A-B MÚLTIPLE (A-B-A-B)

Este diseño es una extensión del diseño A-B-A-B. Desde luego, también se le debe considerar como un diseño de replicación intrasujeto y reversible. Mediante este diseño se puede tener una sucesión de réplicas de la secuencia básica observación-tratamiento (A-B).

No se suele encontrar en la literatura trabajos que reporten el uso de este diseño. La razón fundamental, en el campo de la investigación aplicada (ACA), es que, una vez establecida la relación funcional (fase A_1-B_1) y ya que ha sido apropiadamente replicada (fases A_2-B_2), generalmente no hay justificación alguna para intentar réplicas posteriores. Recuérdese que la investigación aplicada trata de resolver problemas, más que demostrar ciertos principios científicos.

En el campo de la investigación básica, las explicaciones plausibles parecen ser: *a*) aunque se efectúen numerosos experimentos en los que se "activa y desactiva" la conducta del sujeto (secuencias A-B-A-B) durante múltiples ocasiones, el investigador podría considerar innecesario publicar toda la secuencia; *b*) ya que ha logrado una exitosa replicación directa (A-B-A-B), el investigador podría entonces intentar una replicación sistemática, en vez de ampliar su diseño. Esto le permitiría demostrar la solidez de sus datos y un mayor grado de generalización, o *c*) después de la primera replicación (A_2-B_2), el experimentador podría tratar una VI diferente, cambiando así su diseño (por ejemplo, a uno A-B-A-B-A-C).

Un ejemplo de replicación intrasujeto múltiple (diseño A-B múltiple) se encuentra en el libro de Ferster y Skinner (1950). Al tratar los efectos del tiempo-fuera de reforzamiento, estos investigadores indican que en ocasiones el TF (tiempo-fuera) puede producir pausas en la conducta diferentes de las producidas inmediatamente después del reforzamiento en un programa de razón fija. Presentan algunos registros acumulativos (págs. 122-123), en los cuales se alternan los procedimientos de un programa de razón fija (RF) y los de TF. La conducta del animal varía, correspondiendo a los cambios de procedimiento. En este ejemplo (aunque Ferster y Skinner no mencionan el diseño) se puede considerar como fase A al programa de RF y como fase B al TF. Glynn (1970) utilizó un diseño A-B-A-B-A-B para probar la efectividad de diferentes formas de administrar reforzamiento con fichas en un escenario escolar.

Otras investigaciones en las que se empleó alguna forma de este diseño son las de Terrace (1966), Mann (1972), Holland y Skinner (1970) y Herbert y colaboradores (1973).

El principal propósito al usar este diseño es demostrar, a discreción del investigador, que se ha logrado alcanzar un alto grado de control sobre la conducta de determinado organismo. Se obtiene una línea base (por ejemplo, bajo un programa de intervalo variable); durante la fase A se incluye la VI, produciendo un marcado cambio en la línea base (fase B), se retira la VI (vuelta a la fase A), se reincluye la VI (B), se retira (A), se reincluye (B), y así sucesivamente.

Para evaluar este diseño, se aplican todas las consideraciones tratadas en el diseño anterior. Además, debe agregarse que, salvo algunas excepciones, el diseño es prácticamente inaplicable en el ACA; y aun en la investigación básica pudiera ser más productivo explorar los efectos de diferentes valores de la VI durante las fases B.

7.4. DISEÑOS BALANCEADOS CONDUCTUALES

Dentro del AEC hay varias maneras de balancear las condiciones a las que son sometidas las unidades experimentales (sujetos o grupos de sujetos). Una manera, la más usual, fue estudiada en el capítulo 6. A este tipo de diseño balanceado se le conoce, en términos generales, como *diseño balanceado intersujetos* (Underwood, 1972) o *diseño balanceado intergrupos*. Cuando se tienen múltiples condiciones, el diseño balanceado puede tener la forma siguiente:

$$\begin{bmatrix} \text{A-B-A} \\ \text{B-A-B} \end{bmatrix},$$

esta otra

$$\begin{bmatrix} \text{A-B-A-B} \\ \text{B-A-B-A} \end{bmatrix}$$

o alguna extensión de esta última, donde a la primera hilera (A-B-A o A-B-A-B) se le considera como una replicación intrasujeto o intragrupo, con la participación de un solo sujeto (o grupo de sujetos). A la segunda hilera (B-A-B o B-A-B-A) se le considera como otra replicación intrasujeto o intragrupo.

Es importante hacer notar que un diseño que se extiende en dirección horizontal (\rightarrow) indica el uso repetido del mismo sujeto o grupo de sujetos. La contraparte del diseño (es decir, la secuencia que produce el balanceo) también se extiende en la misma dirección, siguiendo las condiciones el orden opuesto. Por la segunda secuencia de condiciones pasan diferentes unidades experimentales.

Otra manera de obtener un balanceo sería:

ABA-BAB

o, en forma similar,

ABAB-BABA

o cualquier extensión en este sentido.

En estos casos se estaría haciendo una replicación directa intrasujeto (o intragrupo), y el guión indica que debe transcurrir un periodo lo suficientemente grande como para poder pasar a la misma unidad experimental por la segunda secuencia. Este periodo debe tener la función de impedir, hasta donde sea posible, los efectos de la última condición de la primera secuencia del diseño que se refleje en la primera condición de la segunda secuencia del diseño. Underwood (1972) considera a este diseño como una técnica para distribuir uniformemente el "error progresivo".

Hay dos razones para preferir al primer tipo de diseño, aunque la primera de ellas concierne a todos los diseños balanceados reversibles: *a*) se reconoce generalmente (Sidman, 1960; Hope, 1970; Underwood, 1972) que ningún diseño balanceado, realmente cancela los efectos de una VI sobre una VD, sino que simplemente facilita la apreciación de los efectos del orden en que se dan los tratamientos sobre el comportamiento de la VD; *b*) es preferible utilizar organismos totalmente independientes en las distintas secuencias de las condiciones, ya que los dos aspectos principales que se trata de conocer mediante este diseño (la relación funcional entre VI y VD y los efectos del orden sobre la VD) se aprecian más claramente con diferentes sujetos. Así, se elimina la posibilidad de que los efectos a largo plazo de la primera secuencia confundan la interpretación de los datos. Azrin y sus colaboradores (1968) aplicaron un diseño balanceado ABA-BAB en algunos de sus sujetos, en un trabajo sobre control postural.

7.5. DISEÑO DE LÍNEA BASE MÚLTIPLE

Kazdin y Bootzin (1972) consideran a este diseño como una alternativa razonable al diseño reversible. La misma posición adoptan Baer y sus colaboradores (1968). Para ellos, esta alternativa es particularmente valiosa cuando se supone que una conducta es irreversible o cuando se estima que la reversión es indeseable. Cabe añadir que este diseño es, de hecho, más eficaz que el diseño reversible. Con él se puede demostrar la relación causal entre las VIs y las VDs en una forma tan convincente como mediante el diseño reversible. Asimismo, el control experimental que se ejerce sobre la conducta puede ser tan contundente como el que se ejerce con aquel diseño. La mayor eficacia del diseño de línea base múltiple radica en que se pueden estudiar varias respuestas simultáneamente, lo cual no se puede hacer con líneas base únicas. Además, siempre queda la posibilidad de utilizar un diseño de línea base múltiple y también hacer *reversiones*. Esta posibilidad excede las justificaciones expresadas por Baer y sus colaboradores para el uso

de este diseño. De hecho, ya se han reportado estudios en los que ambos diseños se han utilizado en combinación (véase por ejemplo, Hall y colaboradores, 1970; Pinkston y colaboradores, 1973).

Hall y sus colaboradores (1970) analizaron tres tipos principales de diseños de línea base múltiple: *a*) respuestas, que es la forma estudiada por Baer y colaboradores; *b*) a través de individuos, y *c*) a través de situaciones. En el primer tipo de línea base múltiple se obtiene un número aceptable de observaciones en diferentes respuestas (de acuerdo al criterio de estabilidad), que sirven como nivel de comparación para evaluar los cambios producidos por la VI. Entonces, se procede a aplicar la VI a cada una de las respuestas (VD) a la vez. Si se observa un cambio en la respuesta en cuestión mientras las demás permanecen inalteradas (o muestran cambios mínimos), esto se interpreta como una relación causal entre la VI y la VD. Entonces se procede a hacer lo mismo con otra de las respuestas, y así sucesivamente. Si después de cada aplicación de la VI a una nueva VD se observa un cambio apropiado, entonces se puede afirmar que la VI es efectiva y en qué grado. Esto, para Baer y sus colaboradores, constituye una forma clara de replicación.

El segundo tipo de diseño de línea base múltiple consiste en obtener líneas base (de una sola conducta) de diferentes individuos y en aplicar, entonces, la misma VI sucesivamente a cada uno de los sujetos a la vez. De nuevo se espera que si la VI es efectiva, se notará un cambio en la línea base de cada individuo como consecuencia de la presencia del tratamiento experimental, y no en ningún otro momento.

El tercer tipo de línea base múltiple (a través de situaciones) requiere que se registre la misma línea base del mismo sujeto en diferentes situaciones (distintos lugares, horas, personas, etc.). Una vez obtenidas las líneas base, se procede a aplicar el procedimiento experimental en cada una de las diferentes situaciones, exclusivamente. La conexión entre la VI y las VDs se puede establecer de acuerdo a los cambios observados en éstas como consecuencia de la aplicación de aquélla.

Fixsen y sus colaboradores (1972) utilizaron el diseño de línea base múltiple al entrenar a seis niños a informar de su propia conducta y de la de sus compañeros. Fjellstedt y Sulzer-Azaroff (1972) emplearon un sistema de fichas y el diseño de línea base múltiple, a fin de modificar la latencia de respuesta de un niño. En el trabajo de Christopherson y colaboradores (1972), estudiado en el capítulo 6, también se utilizó este diseño.

El diseño también se ha empleado con diversas poblaciones y para diferentes problemas (Cooper y colaboradores, 1970, con maestros, a fin de incrementar su atención a conductas apropiadas; Panyan y colabo-

radores, 1970, con personal no profesional, para usar entrenamiento operante; Mandelker y colaboradores, 1970, con maestras de preprimaria, para medir su atención a las conductas apropiadas de los niños). Se ha usado con retardados (Barton y colaboradores, 1970), estudiantes con bajo rendimiento (Hall y colaboradores, 1970), con niños normales (Broden y colaboradores, 1971; Pinkston, 1973), con niños predelinquentes (Bailey y colaboradores, 1970) y con esquizofrénicos paranoides (Lieberman y colaboradores, 1973). En este mismo libro se encuentran ilustraciones de líneas base múltiples en las figuras 6, 9, 10, 11, 12, 17, 18 y 21.

Kazdin y Bootzin (1972) señalaron una posible desventaja del diseño de línea base múltiple a través de respuestas: la generalización de la respuesta. Ésta se puede producir, a pesar de que las respuestas bajo estudio sean aparentemente independientes. También apuntaron que hay una posible deficiencia del diseño cuando se emplea la versión "a través de individuos", cuando se altera la conducta de un sujeto; se puede afectar la ejecución de otros sujetos, a pesar de que todavía estén en línea base. Finalmente, cabe indicar que el diseño de línea base múltiple, en general, demanda mayores recursos (observadores, por ejemplo), que diseños más simples, y esto lo vuelve poco accesible.

II. CLASE TRADICIONAL

7.6. DISEÑO DE MÁS DE DOS GRUPOS ALEATORIOS

Aunque en términos formales este diseño es simplemente una extensión del diseño de *dos grupos aleatorios*, hay ciertas diferencias en cuanto al poder (cantidad y tipo de información) y la economía (costo monetario, esfuerzo, tiempo, etc.), entre ambos diseños.

Al comparar este diseño, que para abreviar denominaremos *diseño multigrupos* (siguiendo a McGuigan, 1972), con un diseño de dos grupos, encontramos las siguientes *ventajas*:

- a) Podemos hacer un mayor número de observaciones y, por ende, un número considerablemente mayor de comparaciones.
- b) Podemos estudiar un mayor rango de la VI.
- c) Podemos especificar mejor la forma de la relación funcional entre la VI y la VD.
- d) Podemos estudiar con mayor detalle los valores intermedios de la VI.
- e) Al hacer una investigación exploratoria, en cierto número de casos podemos obtener datos más precisos con sólo tres grupos, sin in-

crementar considerablemente el costo de la investigación en relación a un diseño de dos grupos.

Asimismo, encontramos las siguientes *desventajas*:

a) Es más costoso, en términos del número de sujetos requeridos; se necesita una mayor cantidad de equipo y materiales, un mayor número de ayudantes y, desde luego, mayor presupuesto.

b) Es más demandante (en lo que se refiere al esfuerzo del experimentador) a mantener la constancia de las condiciones, al tiempo que se debe invertir y al que se debe tratar de reducir la probabilidad cada vez más grande (conforme aumenta N) de que ocurran eventos no deseables (VE). Para una comparación más detallada, se remite al lector a la obra de McGuigan (1972) o a la de Plutchik (1968).

Un ejemplo de trabajo realizado con este diseño lo encontramos en Whittaker (1972). Este autor reproduce el trabajo de Boe y Church, quienes asignaron 60 ratas a seis diferentes grupos que recibieron distintos voltajes de choque eléctrico (VI): 0, 35, 50, 75, 120 y 220. Estos animales recibieron previamente reforzamiento continuo como consecuencia de la respuesta de accionar una palanca. Se observó que a medida que aumentaba la cantidad de voltaje, tanto durante la etapa de castigo como en la de extinción, disminuía la cantidad de respuestas. Se encontró, también, que una función exponencial negativa encajaba adecuadamente con los datos observados.

Underwood (1973) empleó un diseño de tres grupos aleatorios en un estudio de la atención selectiva y la interferencia sobre el procesamiento en memoria. Barber y Folkard (1972) también usaron un diseño de tres grupos en un trabajo acerca de los efectos que ejerce la incertidumbre del estímulo en el tiempo de reacción. Restle y Emmerich (1966) emplearon un diseño de cuatro grupos aleatorios en uno de sus experimentos sobre los efectos del número de problemas sobre la obtención de conceptos. Kendler y sus colaboradores (1972) asignaron sujetos a cuatro, seis y tres condiciones experimentales en experimentos sucesivos. Hay y Goldsmith (1973) emplearon un diseño de cuatro grupos aleatorios en un experimento sobre percepción del movimiento.

A continuación compararemos el diseño multigrupos con el diseño intrasujeto (reversible). De primera intención parecería que representa lo mismo someter a un mismo sujeto a los diferentes valores de la VI, que hacer que distintos grupos de sujetos sean sometidos exclusivamente a cada uno de ellos. Esto nos llevaría de nuevo a la situación examinada en la sección 4.4.

Como se vio en dicha sección, en términos generales la aproximación de grupo solamente da una descripción idealizada del comportamiento de un "sujeto promedio". En el caso específico del diseño multigrupos,

éste puede dar una indicación de la forma ideal de la relación funcional entre la VI y la VD. Lo más probable es que ningún sujeto particular siga con exactitud la relación funcional ideal, y es posible que alguno se desvíe notablemente de ella. Sidman (1960) va más lejos y afirma que varios autores han demostrado que este tipo de diseño (multigrupos) puede producir una descripción distorsionada (y hasta falsa) de la conducta de un organismo individual.

Cuando se utiliza el diseño multigrupos para indicar la forma de la relación funcional, ésta va a depender fundamentalmente del número de sujetos en cada condición y de la variabilidad presente en cada grupo. Por tanto, se podrían obtener diferentes variantes de la función dependiendo de dichos parámetros, y, en casos extremos, incluso distintas funciones.

Aunque se puede obtener una indicación de la relación funcional, el diseño no permite saber qué es lo que sucede en la VD durante la transición de un valor de la VI al siguiente.

Se puede concluir que, por una parte, este diseño es más poderoso que uno de dos grupos, aunque es más costoso; por otra, se le puede considerar como un pobre sustituto de un diseño intrasujeto cuando se trata de describir la conducta de organismos individuales como consecuencia de cambios en la VI. Por otra parte, puede ser una herramienta valiosa cuando se trata de sintetizar el comportamiento de grupos de individuos ante diferentes valores de la VI, y no se debe descartar su uso en combinación con la metodología de la línea base. Además, el diseño puede proporcionar información útil cuando no es posible someter a la misma unidad experimental a las diversas condiciones (por falta de tiempo o control sobre ellas, por ejemplo).

7.7. DISEÑOS CONTRABALANCEADOS

Hay un acuerdo general en dividir estos diseños en dos clases: *completos* e *incompletos*. Este tipo de diseño no debe confundirse con los diseños balanceados del AEC.

Los diseños contrabalanceados son aquellos en los cuales cada grupo de sujetos (o, a veces, cada sujeto individual) es sometido a todas las condiciones o valores de la VI, con las siguientes restricciones: *a*) cada unidad experimental sólo es sometida una vez a cada condición; *b*) diferentes grupos (o sujetos) pasan por dichas condiciones en diferentes órdenes, y *c*) los grupos (o sujetos) deben asignarse al azar a las diferentes secuencias. Además, debe agregarse que la tecnología

de la línea base no se utiliza y que no hay replicaciones intragrupos ni intrasujetos.

Un ejemplo de un diseño contrabalanceado completo con tres valores de la VI se presenta en la tabla 1.

Tabla 1.

UN DISEÑO CONTRABALANCEADO COMPLETO

		<i>Valores de la VI</i>		
Unidad experimental	1	1	2	3
Unidad experimental	2	1	3	2
Unidad experimental	3	2	1	3
Unidad experimental	4	2	3	1
Unidad experimental	5	3	1	2
Unidad experimental	6	3	2	1

El número de secuencias posibles coincide con el factorial de los n valores de la VI, que se simboliza como $n!$. Recuérdese que este número $n!$ se obtiene multiplicando sucesivamente todos los números comprendidos entre 1 y n (en cualquier orden). Así, en el caso de $n=3$, $n!$ sería igual a $1 \times 2 \times 3 = 6$, que es el número de secuencias que se obtuvieron en el ejemplo anterior. Para $n=4$, el factorial sería igual a $1 \times 2 \times 3 \times 4 = 24$, y así sucesivamente.

Un ejemplo de diseño contrabalanceado incompleto con tres valores de la VI se ilustra en la tabla 2.

Tabla 2.

UN DISEÑO CONTRABALANCEADO INCOMPLETO

		<i>Valores de la VI</i>		
Unidad experimental	1	1	2	3
Unidad experimental	2	2	3	1
Unidad experimental	3	3	1	2

Como se ve, en ambos casos cada valor de la VI ocurre el mismo número de veces en cada posición ordinal; cada uno se ve precedido de los restantes el mismo número de veces; cada unidad experimental solamente pasa por cada condición una sola vez, y cada unidad expe-

rimental recibe una secuencia diferente. La principal diferencia es que, en el diseño incompleto, no se presentan todas las combinaciones posibles de las condiciones, como sucede con el diseño completo.

El objetivo de los diseños contrabalanceados es doble: *a*) evaluar los efectos residuales que produce un tratamiento en el siguiente, y *b*) evaluar los efectos del orden en que se presentan las condiciones experimentales. Al igual que con otro tipo de diseños que tienen que ocuparse de este problema, este tipo de diseño no controla o cancela dichos efectos, sino que sólo permite que se les aprecie objetivamente.

Las ventajas de este tipo de diseño sobre un diseño más simple, como uno de grupos apareados o uno de grupos aleatorios, son obvias: además, de proporcionar cuando menos la misma información que aquéllos, permite la posibilidad de evaluar los efectos del orden y los efectos residuales. Por otra parte, también presenta sus desventajas; todas se relacionan con "crecer" la forma desmedida en que un diseño de esta clase puede incrementarse a medida que aumentan los valores de la VI. Las tres principales son: *a*) el número de sujetos requeridos puede ser "estratosférico", sobre todo en los diseños completos y si se usa con grupos como unidades experimentales. [D'Amato (1970) demostró que el número de sujetos debe ser un múltiplo del número de condiciones; y si se ha de cumplir con las restricciones estadísticas, este múltiplo no puede ser uno de los mínimos]; *b*) cuando el número de valores de la VI es grande se puede llegar al caso de tener solamente un sujeto por grupo, lo cual aumentaría la variabilidad entre las condiciones de una manera no deseable (el diseño incompleto presenta la ventaja de minimizar ambos problemas), y *c*) cuando hay unos cuantos sujetos por condición, la variabilidad dentro de cada condición puede ser demasiado grande.

Cuando se compara esta clase de diseño con los diseños balanceados conductuales, se les encuentra en clara desventaja; por una parte, la confiabilidad de aquéllos, sobre todo cuando presenten algún tipo de replicación intrasujeto, es mayor. También, con un número menor de secuencias es posible obtener tanta información como mediante un diseño contrabalanceado. Además, debe recordarse que un diseño balanceado también permite evaluar los efectos de orden y los residuales. La demanda del número de sujetos se reduce considerablemente y el peligro de una gran variabilidad dentro de las condiciones se controla absolutamente con el uso de una línea base adecuada. Comoquiera que sea, esta es una clase útil e interesante de diseños, sobre todo para el investigador que dispone de un gran número de unidades experimentales y durante cortos periodos. Para una exposición más detallada de este tipo de diseños, consúltese el trabajo de D'Amato (1970).

8

diseños multivariables

LOS DISEÑOS QUE se estudiarán en este capítulo son el producto de situaciones experimentales considerablemente más complejas que las analizadas en capítulos anteriores. Junto con la complejidad de los experimentos aumenta también el poder de los diseños. Los diseños multivariables pueden contestar preguntas de investigación que no se podrían responder con diseños más simples y, en ocasiones, ni siquiera con combinaciones de éstos.

El término *multivariable* se refiere, en este contexto, a que en esta clase de diseños es posible la *manipulación de un número igual o mayor a dos VI*. No se le debe confundir con el empleo que se le da en estadística o en la teoría de la medición, donde se le usa para describir el tratamiento de datos producidos por la *medición de múltiples VD*.

Respecto al número de condiciones, se considera redundante indicar que los diseños multivariables son multicondicionales. Obviamente, no puede haber un diseño multivariable que sea bicondicional; por tanto, se ha omitido el calificativo correspondiente a la cantidad de condiciones.

Una de las características principales de los diseños multivariables, además de prácticamente permitir el estudio de cualquier rango de las VI, es la de permitir también el análisis de los efectos combinados

de dos o más VI sobre la(s) VD. Permiten, asimismo, separar estos efectos conjuntos de los efectos de cada VI por separado.

I. CONDUCTUAL

8.1. DISEÑOS REVERSIBLES MULTIVARIABLES

Bajo este rubro se incluyen aquellos diseños que reúnen las siguientes características: *a)* presentan por lo menos una reversión, es decir, hay por lo menos una recuperación de la línea base original; *b)* al usarlos se manipula más de una VI, ya sea una a la vez o varias de ellas conjuntamente; *c)* el orden de las secuencias puede variar, es decir, un diseño pudiera ser A-B-A-C o A-B-C-A; *d)* no hay una restricción en cuanto al número de condiciones; un diseño multivariable pudiera ser tan largo como A-B-C...X-Y-Z-A.

La denominación de este tipo de diseño podría confundirse con el del diseño A-B múltiple. Recuérdese que este último es univariable, y aunque en él se pueden probar *diferentes valores de la misma VI*, sólo hay una VI que se denomina genéricamente con una "B". Aun cuando un diseño A-B múltiple puede tener un gran número de reversiones (por ejemplo: A-B-A-B-A-B-A...B) nunca incluirá una segunda o tercera VI (que se designaría con una letra diferente de la B).

En los diseños reversibles multivariables, la característica principal es la eventual recuperación de la línea base, ya que ésta es fundamental para hacer comparaciones y para contrarrestar los efectos residuales hasta donde fuese posible.

Veamos ahora algunos ejemplos de diseños reversibles multivariables. Hall y sus colaboradores (1972) presentaron una serie de estudios dirigidos por profesionales y realizados por padres de familia que modificaron "conductas problema" de sus hijos. En dos de estos estudios se usó un diseño reversible multivariable. En el primero de ellos se trató de incrementar el porcentaje de tiempo que un niño usaba una pieza ortodóncica (dental). Las fases del diseño fueron:

A₁) Línea base. Durante ocho días se registró intermitentemente si el niño estaba usando el aparato o no. A éste no se le informó del propósito de los registros. Durante esta fase usó la pieza el 25% del tiempo de registro.

B₁) Reforzamiento social. Durante los siguientes nueve días la madre del niño alabó verbalmente a éste cuando al hacer el registro lo encontraba usando el aparato (reforzamiento social contingente a la respuesta). La cantidad de tiempo aumentó al 36%.

C₁) Pago monetario demorado. Durante los siguientes quince días se intentó el uso de una nueva VI. La madre le indicó al niño que por cada vez que lo registrara y que él estuviese usando el aparato dental, ganaría 25 centavos. Si no lo estaba usando, perdería también 25 centavos. El dinero ganado no se pagaba inmediatamente, sino que se anotaba en un calendario y le era pagado al terminar el mes. Durante esta fase la cantidad promedio de tiempo aumentó al 60%.

D₁) Pago monetario inmediato. En esta fase, que duró 18 días, el pago fue hecho inmediatamente después de cada registro (eran cinco registros diarios). Ésta fue la tercera VI. La media de tiempo aumentó al 97%.

A₂) Línea base. En esta fase se intentó la recuperación de la línea base original. Al niño se le informó que, en vista de que su conducta había producido un mejoramiento en su estructura vocal, el intercambio monetario dejaba de ser necesario. Durante esta fase la media de tiempo disminuyó al 64%.

D₂) Pago monetario inmediato. Durante los siguientes 13 días se reincluyó la condición de pago inmediato. La media aumentó al 98.5%.

D₃) Pago monetario inmediato con registro ocasional. Durante esta fase se mantuvieron las condiciones de intercambio monetario, excepto que se le avisó al niño que los registros sólo se harían de cuando en cuando. Durante estos "chequeos" la media de tiempo aumentó al 100%.

En este estudio hay dos aspectos sobresalientes: *a)* la efectividad de la tercera VI (pago monetario inmediato) sobre la VD, y *b)* no se pudo recuperar la línea base original. Esto sugiere la presencia de efectos residuales de los tratamientos previos a la reversión y, fundamentalmente, de la tercera VI.

En el segundo estudio se utilizó otro diseño, el sujeto fue diferente y las conductas a incrementar fueron de tipo casero. El diseño tuvo las condiciones A-B-C-D-A-E, a saber: línea base, sistema de puntos, centavos, uniforme de campamento, línea base y regalos de Navidad. Como se ve, en este estudio se utilizaron cuatro VI y se hizo una reversión.

Un estudio, en el cual se manipularon dos VI simultáneamente en un diseño reversible multivariable, fue el de Medland y Stachnik (1972). Estos investigadores aplicaron la técnica llamada "el juego de buena conducta", a fin de disminuir la frecuencia de "conductas indeseables" en un salón de clases. Dichas conductas eran, básicamente, tanto de tipo destructivo, como de hablar y levantarse de su lugar.

El diseño experimental constó de seis fases: A-B-A-C-D-B. Durante las fases A se registró la frecuencia con que ocurrían las conductas objetivo en el transcurso de la clase de lectura. Se había dividido a los estudiantes en dos equipos de 14 miembros cada uno. Esta fase duró cinco sesiones. Durante las fases B se incluyó la primera VI, que consistía en el "juego de buena conducta". Básicamente, dicho juego consiste en indicar a los estudiantes cuáles respuestas deben eliminar (las conductas objetivo "indeseables"), evaluando su ocurrencia en términos de los equipos y no de los sujetos individuales. Ambos equipos podían "ganar" y el reforzamiento consistía en recibir tiempo extra de recreo como consecuencia de la "buena conducta". Cuando las conductas eran las "apropiadas" se encendía una luz verde, y cuando eran "inapropiadas", una luz roja. La fase C consistió en la repetición de las "reglas de clase" de parte del maestro; durante esta fase no había consecuencias reforzantes. Finalmente, la fase D consistió en la combinación de una de las VI (las reglas) con parte de otra VI (las luces). La figura 23 presenta registros individuales de algunos de los estudiantes más destructivos durante la primera fase A. Como se puede ver, la técnica del "juego de buena conducta" (fases B) fue la más efectiva para la reducción de las conductas objetivo. También se puede observar que durante la reversión (fase A₂) hubo recuperaciones de las líneas base originales. Finalmente, se puede observar que, una vez establecido el efecto del "juego" (y de sus consecuencias), se le puede sustituir por una combinación de las luces y las reglas, a fin de mantener las conductas "inapropiadas" a un nivel notablemente bajo.

Otro estudio reciente en el cual se utilizó un diseño reversible multivariable fue realizado por O'Brien y Azrin (1972). Estos investigadores modificaron la conducta de ingerir alimentos (comer "apropiadamente") en una niña con retardo en el desarrollo. Se empleó un diseño A-B-A-C-D-A-B-A con las siguientes etapas específicas: *a*) las etapas A comprendieron la toma de una línea base observacional; *b*) las etapas B consistieron en lo que O'Brien y Azrin denominaron "interrupción-extinción". Durante esta etapa se impedía que la niña continuara comiendo; para ello, se le detenía la mano. Si la niña tomaba la comida en sus manos, éstas se le limpiaban con papel toalla; *c*) la etapa C consistió en dirección manual; ésta se refería a que la maestra le colocaba la cuchara en la mano, le tomaba ésta y la obligaba a que se la llevara a la boca, etc., y *d*) la etapa D consistió en la combinación de los procedimientos de las etapas B (interrupción-extinción) y C (dirección manual).

Como se puede observar en la figura 24, la combinación de la VI (interrupción-extinción) con la VI (dirección manual) fue la más

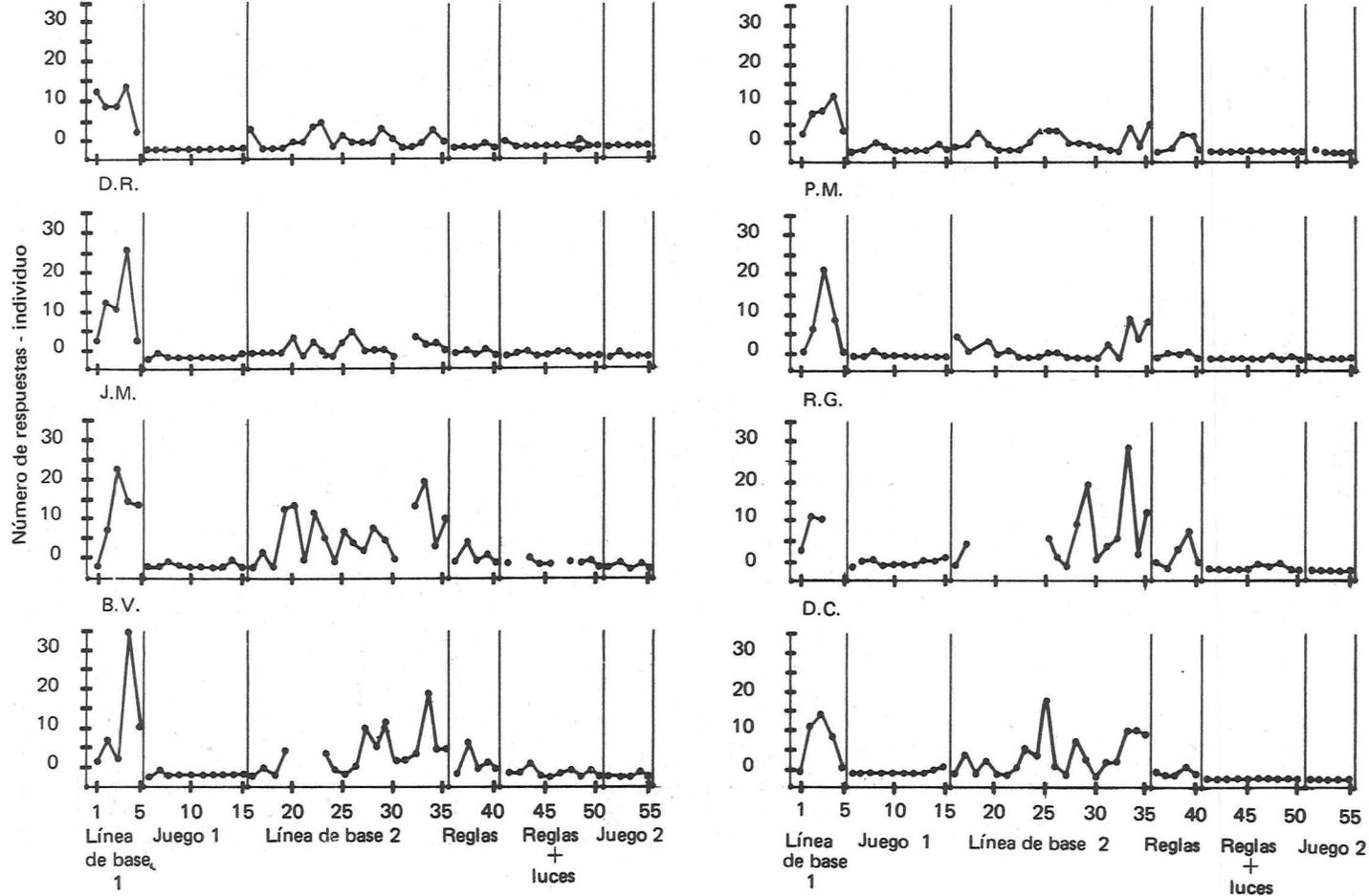


Figura 23. Gráficas que muestran los registros de cuatro estudiantes que emitieron el número más grande de conductas objetivo en cada grupo durante la fase de línea base₁. Las conductas de platicar, de estar fuera de su lugar y destructivas se han combinado en las gráficas individuales. (Tomado y adaptado de Medland y Stachnik, *Journal of Applied Behavior Analysis*.)

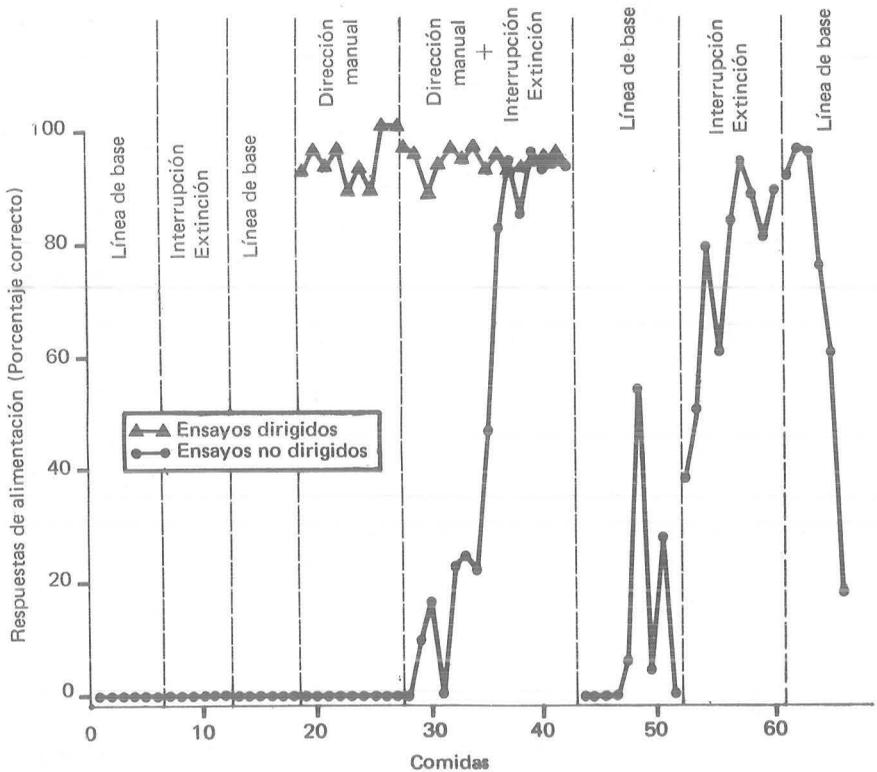


Figura 24. Porcentaje de respuestas de alimentación completadas correctamente por una niña de 6 años retardada profunda. Se le dieron cinco comidas ligeras durante un periodo de 6 horas cada día. Durante el procedimiento de "interrupción-extinción" la maestra impedía que la niña comiera, quitándole la comida de la mano durante una respuesta incorrecta. Durante el procedimiento de dirección manual, la maestra ayudó a la niña en los ensayos dirigidos, por medio de la dirección manual del movimiento apropiado. Durante la línea base no se usó interrupción-extinción ni dirección manual. (Tomado y adaptado de O'Brien y Azrin, *Journal of Applied Behavior Analysis*, vol. 5, núm. 4, 1972, págs. 389-399. Copyright 1972 por la Society for the Experimental Analysis of Behavior, Inc.)

efectiva para incrementar considerablemente el porcentaje de respuestas "correctas".

Hay dos aspectos de este estudio que son sobresalientes desde el punto de vista del diseño experimental: *a)* el hecho de que en este diseño se incluyeron varias reversiones, y *b)* el hecho de que se estu-

diaron los efectos separados de dos diferentes VIs y también sus efectos simultáneos. Cabe hacer notar, también, que una vez que se estableció un cambio de la conducta a través de la interacción de las dos VIs, sólo fue necesario utilizar una de ellas para mantener un alto nivel de "respuestas apropiadas".

Finalmente, se puede observar que el tener más de una reversión en un diseño sirve como una forma de control que permite una clara evaluación de los cambios contingentes a la inclusión de una VI, así como de sus efectos residuales.

II. CLASE TRADICIONAL

8.2. DISEÑOS FACTORIALES

Si un investigador deseara obtener una indicación del efecto de una VI sobre una VD, podría usar *un diseño de dos grupos aleatorios*, a fin de aclarar su duda. Si, además, quisiera tener una indicación del efecto de una segunda VI sobre la misma VD, entonces podría utilizar *otro diseño de dos grupos aleatorios* y, seguidamente, distintos sujetos. Aunque con este procedimiento podría obtener ambas indicaciones, hay una pregunta que no podría contestar con esta clase de diseño: ¿Cuál es el efecto conjunto de ambas VI sobre la misma VD?

Un diseño factorial sí es capaz de contestar esta pregunta, además de las de los efectos simples de las VI por separado (en la terminología de los diseños factoriales, se denomina a dichos efectos como "efectos principales"). Al contestar esta clase de pregunta, el diseño factorial no requiere que se aumente el número de grupos. En la figura 25 se compara la solución de dos diseños aleatorios con la de un solo diseño factorial.

El diseño que se muestra en la figura 25 se conoce como *diseño factorial de 2×2* . Cada número (independientemente de su valor), separado por un símbolo de multiplicación, indica la presencia de una VI o de un factor en la jergonza de los diseños factoriales. El valor que asume dicho número indica la cantidad de valores de la VI en cuestión. Además, al multiplicar los números implicados en un diseño factorial, se obtiene el número de condiciones experimentales (generalmente grupos), requeridas por el diseño. Así, un diseño factorial de 2×3 tiene 2 VI; la primera de ellas asume dos valores y la segunda tres. El número de grupos (o condiciones) requeridos es seis (véase la figura 26 para una ilustración).

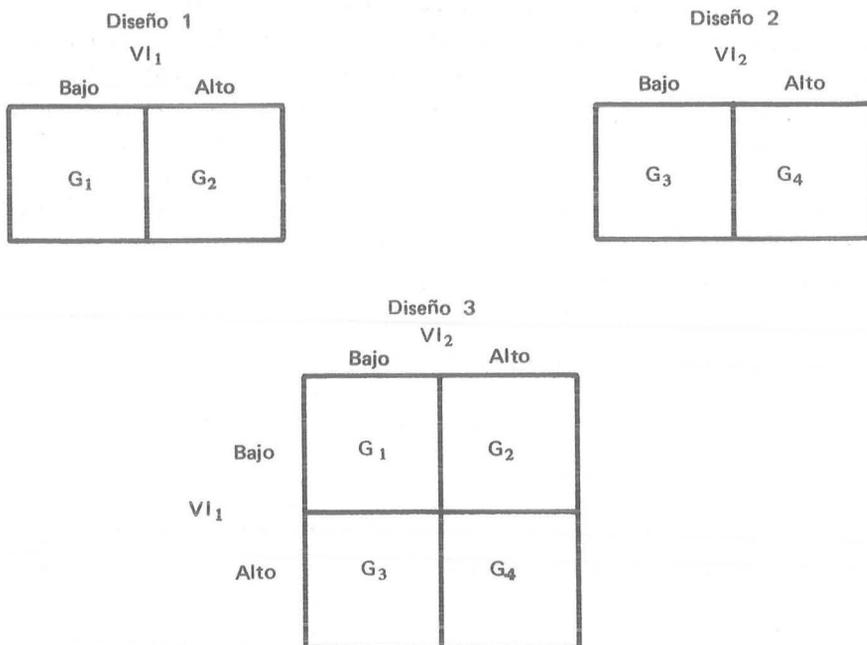


Figura 25. En la parte superior se muestra un esquema con dos diseños de dos grupos aleatorios separados. Cada VI asume dos valores y cada grupo asume sólo un valor de una VI. En la parte inferior se muestra un diseño factorial con dos VI, que asumen dos valores cada una, y cada grupo recibe un valor de cada VI.

Un diseño factorial de $2 \times 2 \times 2$ tiene 3 VI (o factores) con dos valores cada uno y un total de ocho grupos. Su esquema se presenta en la figura 27.

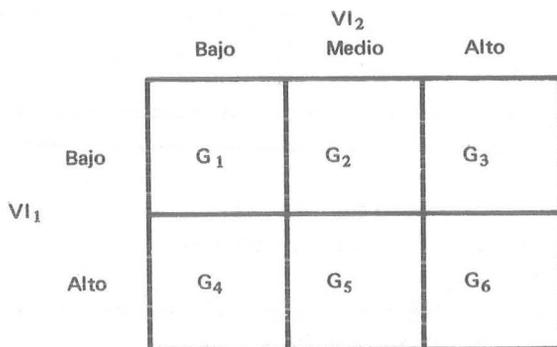


Figura 26. Un diseño factorial de 2×3 . La primera VI tiene dos valores, la segunda tres y se forman seis grupos.

Existen medios estadísticos y gráficos para analizar los datos obtenidos con los diseños factoriales. Estos últimos, sin embargo, son menos comprensibles a medida que los diseños aumentan en complejidad.

El medio estadístico para analizar los diseños factoriales (así como otros diseños más simples o más complejos) es el AVAR y quizá debido a la gran frecuencia con que se utiliza, algunos autores han descrito erróneamente ciertos experimentos implicando el uso de “un AVAR de 2×2 ” o de “un AVAR de 3×4 ”, etc.

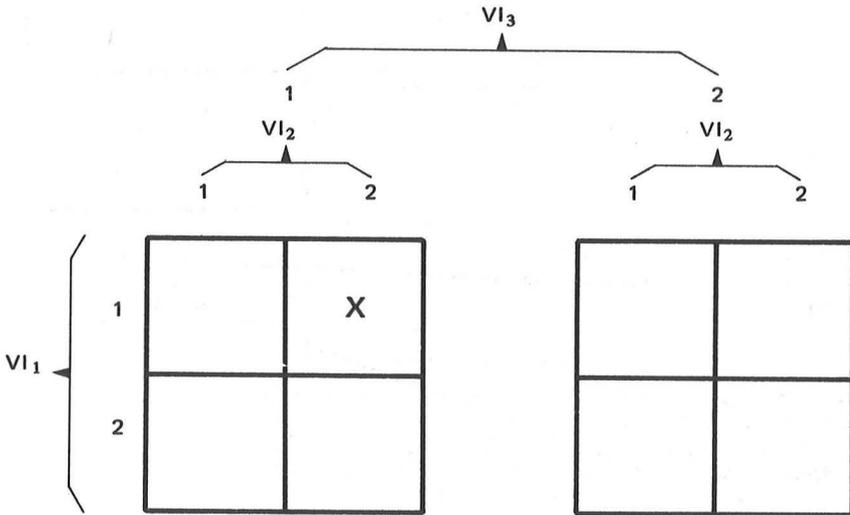


Figura 27. Un diseño factorial $2 \times 2 \times 2$. La celdilla marcada con una X corresponde a un grupo de sujetos que recibiría el valor 1 de la VI_1 , el valor 2 de la VI_2 y el valor 1 de la VI_3 .

Sin entrar en detalles, se puede decir que el AVAR produce ciertas cantidades (razones F) que indican qué tan importante es el efecto de una o más VI sobre la VD. En general, mientras más se aleje esta cantidad (F), de 1, mayor será el efecto. Para el lector interesado en el estudio del AVAR, le sería útil consultar a Young y Veldman (1972), Hays (1963) o Kelly y colaboradores (1969), quienes se aproximan a esta técnica de formas bastante diferentes entre sí.

Gráficamente también es posible advertir los efectos de la(s) VI sobre la VD. Por lo general, se usan gráficas bidimensionales en las cuales el eje horizontal representa los valores de una VI y el eje vertical indica los valores de la VD. Las líneas que se obtienen representan la forma en que afectan los valores de una segunda VI a la VD, en relación a los valores de la primera VI. La figura 28 muestra una

tabla con datos ficticios de un diseño factorial de 2×2 y una gráfica que relaciona a ambas VI y a la VD.

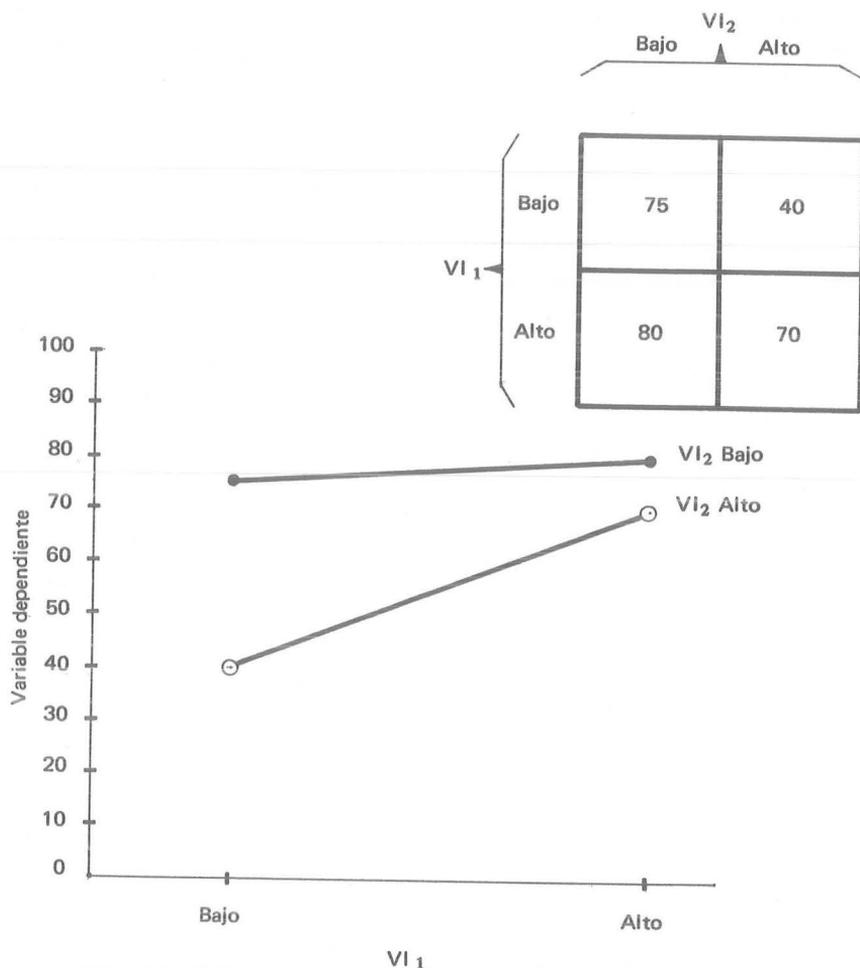


Figura 28. Datos ficticios que muestran el efecto de la interacción de las VI_1 y VI_2 sobre la VD. Los puntos representan las medias aritméticas de las condiciones.

En el caso ilustrado en la figura 29 se pueden notar grandes diferencias entre las puntuaciones de los grupos que recibieron los valores bajos y los que recibieron los valores altos de las dos VI. Probablemente, ambas diferencias corresponden a "efectos principales" notables de cada VI sobre la VD; sin embargo, estos datos demuestran que no

hubo un efecto conjunto de ambas VI sobre la VD. Esto se puede encontrar si se observa que las líneas de la gráfica son paralelas. Este paralelismo significa que hay una diferencia constante (mismo signo) entre los niveles alto y bajo de la VI a medida que aumentan los valores de la VI. Cuando sumamos los valores bajos de una VI (independientemente de su clasificación en la otra VI) y comparamos esta

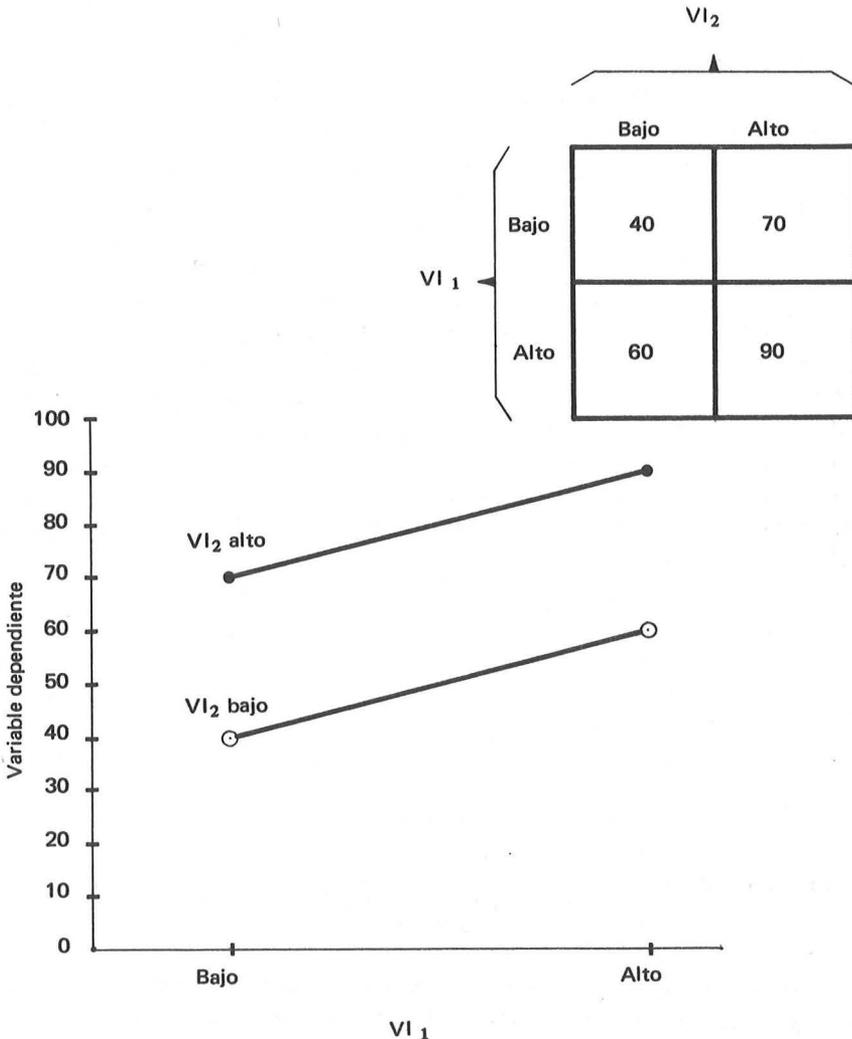


Figura 29. Datos ficticios y gráfica de un diseño factorial de 2×2 que muestran un efecto principal.

suma con la de los valores altos de la misma VI, encontramos una diferencia entre ambas sumas. Esto indica que probablemente hay un efecto principal. Si, en este ejemplo, repitiéramos ese proceso con la otra VI, encontraríamos que probablemente hay otro efecto principal. Ambos son efectos separados de las dos VI; sin embargo, si sumáramos los valores de los grupos extremos (bajo-bajo, más alto-alto) y comparáramos esta suma con la de los valores intermedios (bajo-alto, más alto-bajo), encontraríamos que prácticamente no hay ninguna diferencia entre ambas sumas, es decir, no hay un efecto conjunto o interacción de ambas VI sobre la VD.

En la figura 30 presentamos un caso extremo de ausencia de efectos principales y presencia de una gran interacción.

En este caso se puede notar que las sumas de los dos valores bajos de una VI es igual a la suma de los valores altos. Lo mismo sucede con la otra VI. Se dice entonces que no hay efectos principales. Si estos datos correspondieran a diseños más simples, el investigador se vería tentado a pensar que sus VI no ejercen ningún efecto sobre la VD; sin embargo, si graficamos estos datos, vemos que hay una marcada desviación del paralelismo entre las líneas que representan a los diferentes valores de la VI. De hecho, hay un cruce de líneas sumamente claro que nos indica que *hay un efecto conjunto notable de ambas VI*. A este efecto conjunto se le conoce como *interacción* y constituye la clase de preguntas que no se pueden contestar con diseños más simples. La figura 28 muestra, en contraste, una interacción moderada (nótese que las líneas *no* son paralelas).

Considérese un diseño de $2 \times 2 \times 2$. En este caso, no sólo se pueden contestar las tres preguntas de los efectos separados de las VI que se podrían contestar con diseños más simples, sino que también se pueden contestar preguntas acerca de tres interacciones dobles ($A \times B$, $A \times C$ y $B \times C$), y, además, una pregunta acerca de la interacción triple ($A \times B \times C$). Conforme aumenta el número de VI (factores), se incrementa la cantidad de interacciones, algunas de las cuales llegan a ser realmente intrincadas y en ocasiones carecen de sentido.

Los diseños factoriales se usan ampliamente en la investigación psicológica; basta con revisar un número reciente del *Journal of Experimental Psychology* para constatar este hecho. También se le emplea frecuentemente en la investigación educativa y agrícola. En ocasiones, el diseño se usa midiendo solamente una vez a las unidades experimentales, y en otras ocasiones tomándoles medidas repetidas. En este último caso, es frecuente que cada sujeto reciba ciertos valores de lo que se llama "variable dentro del sujeto" (véanse las secciones 9.3 y 9.4 para una

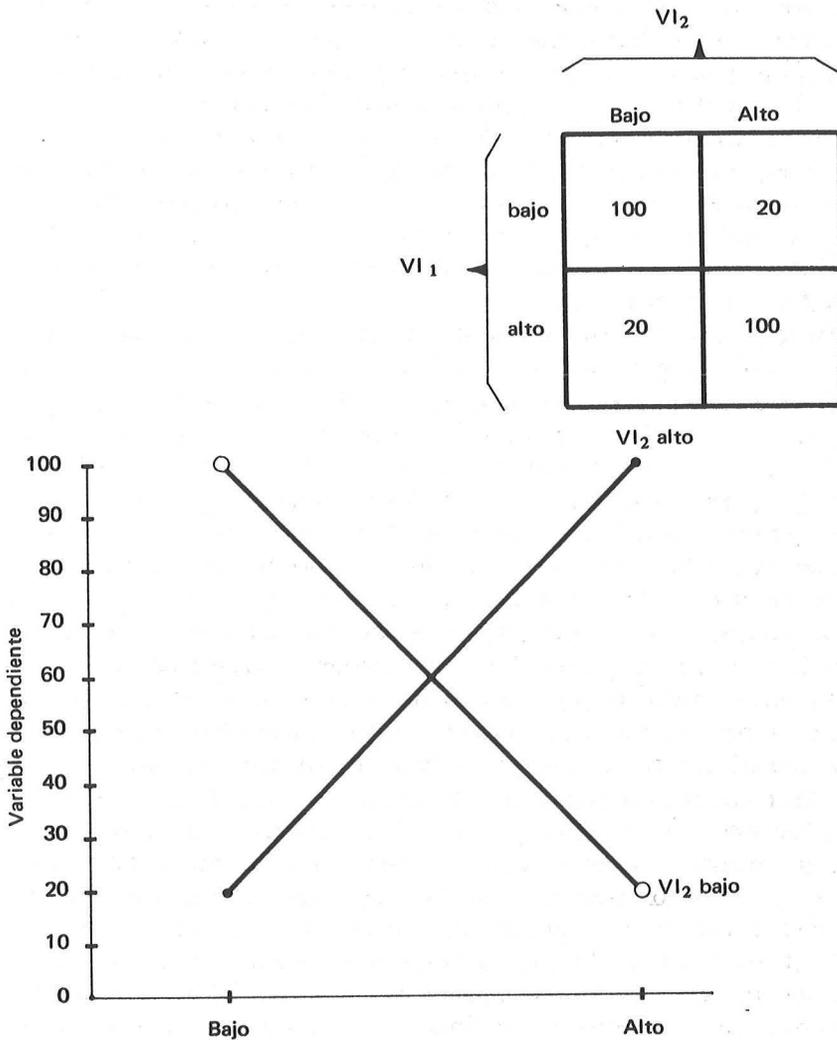


Figura 30. Datos ficticios y gráfica de un diseño factorial de 2×2 que muestran una interacción extrema.

explicación al respecto) y tales valores se mencionan como parte del diseño factorial. Una forma, quizá sobresimplificada, de explicar esta variante del diseño sería la siguiente: cada grupo de sujetos recibe una combinación particular de valores de las “variables entre”, y todos los grupos reciben combinaciones diferentes de dichas variables (son mutuamente exclusivos); por otra parte, cada sujeto, sin importar a

qué condición fue asignado, recibe todos los valores de las "variables dentro", lo mismo que el resto de los sujetos.

Un ejemplo de este uso del diseño lo encontramos en el trabajo de Belleza y Cheney (1973). Estos investigadores emplearon un diseño factorial entre-dentro $4 \times 2 \times 2 \times 4 \times 11$ en el cual sólo la primera VI (forma de una lista) era una variable "entre". Las otras cuatro VI eran dentro del sujeto (condición de aislamiento, intervalo de retención, posición de aislamiento y posición serial). En este estudio se investigó el efecto del recuerdo inmediato y el demorado en tareas relacionadas con la memoria.

De entre la multitud de trabajos que han empleado este diseño hemos muestreado algunos. A continuación los citamos, mencionando qué tipo de factorial se usó. Saunders y Neimark (1967) utilizaron un 2×5 ; Hunt (1961) usó varias veces un 4×4 ; Giambra (1969) un $2 \times 2 \times 2 \times 10$; Briggs y Svanson (1970) un $2 \times 3 \times 3 \times 3$; Santa y Ranken (1972) un $2 \times 8 \times 3$ con medidas repetidas; Grings y Schell (1973) un 2×2 ; Campione (1972) un $2 \times 2 \times 2 \times 2 \times 2$, y Collins y colaboradores (1972) utilizaron diseños factoriales tan complejos como un $6 \times 2 \times 5 \times 6 \times 2$ o un $6 \times 2 \times 2 \times 3 \times 5 \times 2$.

Se pueden construir diseños factoriales que incluyan las principales características de otros diseños más elementales, como el de dos grupos aleatorios, el diseño multigrupos, el de bloques aleatorios y el de mediciones repetidas. Edwards (1968) y especialmente Kirk (1972) analizaron con detalle la construcción y la lógica de tales diseños.

Fisher (1966) demostró que la eficacia de los diseños factoriales no solamente se limita a la posibilidad de investigar las interacciones, sino que mediante técnicas como la de fraccionamiento o la de confusión es posible hacer replicaciones parciales de un experimento. Además, señaló las ventajas que desde el punto de vista económico tiene esta clase de diseños. En su estudio, señaló que los experimentos en agricultura pueden tardar varios meses y que si se usan diseños simples se pueden contestar muy pocas preguntas a la vez, en ocasiones una sola pregunta por varios meses de experimentación. Es obvio que en la ciencia de la conducta y en ciencias afines encontramos el mismo tipo de problema y que los diseños factoriales nos ofrecen el mismo tipo de ventajas económicas que a los agrónomos.

Kish (1972) citó al mismo Fisher, indicando que los diseños factoriales poseen dos ventajas obvias y una sutil sobre los diseños más simples: a) *mayor eficacia*, en el sentido de que con un diseño factorial se obtiene la misma información, requiriéndose solamente una cuarta parte de las observaciones que serían necesarias de otra manera; b) *mayor comprensividad*, en el sentido de que se evalúan las interaccio-

nes además de los efectos principales, y *c*) que cualquier conclusión que se obtenga, tiene una mayor base inductiva, ya que ha sido obtenida a través de la variación de factores, que se hubieran mantenido estrictamente constantes de otra manera. Cochran y Cox (1971) concuerdan en general con esta apreciación, pero hacen notar que en la práctica no todo experimentador tiene los recursos para realizar trabajos que impliquen diseños factoriales con más de dos factores y que, además, a medida que aumenta la complejidad del diseño (aumenta el número de las combinaciones posibles que constituyen los tratamientos) se incrementa el error contenido en los datos.

Al discutir las inconveniencias del diseño, Kirk (1972) coincide en este último punto y además añade: *a*) cuando hay interacciones, el diseño carece de simplicidad en la interpretación de los datos (este punto es particularmente relevante cuando las interacciones son triples o más complejas; de este modo, es muy difícil hallar al presente problemas en la ciencia de la conducta que ameriten siquiera el planteamiento de interacciones de tal naturaleza),* y *b*) el uso de este diseño compromete al investigador a llevar a cabo experimentos relativamente grandes mientras, por otra parte, diseños más sencillos podrían arrojar resultados más atractivos o dirigir su línea de investigación.

* Paréntesis del autor

9

otros diseños

EN VISTA DE que las investigaciones en el campo del diseño experimental abarcan un gran número de diseños, algunos de los cuales tienen cierta relevancia para la ciencia de la conducta y otros son totalmente intrascendentes, en este capítulo se han agrupado los más comunes de la primera categoría. Como la relevancia de cada uno de estos diseños por separado es menor, simplemente se les ha dado la denominación de "otros diseños". Esto no significa que en un estadio más avanzado de la psicología alguno de ellos no pudiese adquirir una mayor relevancia y fuera utilizado con más frecuencia que hasta el presente. El lector interesado en una exposición exhaustiva de diseños experimentales relevantes a ciencias como la biología o la agronomía, puede consultar el libro de Cochran y Cox (1971).

I. CLASE CONDUCTUAL

9.1. DISEÑO REVERSIBLE CON UNA VARIABLE CONTROL

Legítimamente se debería considerar a este diseño como un caso específico del diseño de línea base múltiple. En esta sección se le ha

considerado aparte, por tres razones: *a*) por ayudar a la apreciación de uno de los mayores problemas que se presentan en los diseños reversibles: el de la evaluación de los efectos residuales; *b*) porque permite la estimación cruzada de los efectos de la VI, y *c*) porque es una alternativa más económica que la ejecución de un diseño de línea base múltiple con más de dos líneas base. El diseño parte del fundamento de registrar dos líneas base, una de las cuales será alterada experimentalmente y la otra servirá como control o testigo. Se debe escoger y registrar dos VD que se encuentren lo menos relacionadas posible. Esta independencia entre ambas variables debe estimarse en términos conductuales y no en términos estadísticos. Si las dos respuestas registradas son relativamente independientes y sólo una de ellas se ve afectada por la VI durante la fase B del diseño, el nivel de esta línea base debería cambiar, mientras que el nivel de la variable control debería permanecer aproximadamente igual durante las diferentes fases del diseño reversible. Con el objeto de simplificar esta exposición, se considerará al diseño A-B-A como representante de los diseños reversibles. Suponiendo que las dos variables registradas fueran realmente independientes, entonces se podría estimar la magnitud de los efectos residuales al comparar el nivel A_2 de la variable experimental con el nivel A_1 de la misma y al comparar también los niveles A_1 y A_2 de la variable experimental con los niveles correspondientes de la variable control; la diferencia A_1 experimental - A_1 control debería ser equivalente a la diferencia A_2 experimental - A_2 control. La figura 31 muestra datos ficticios en los cuales se puede apreciar un ligero efecto residual (activo); al empezar la fase A_2 , la variable experimental tiende a disminuir a su nivel original, pero se encuentra mucho más separada del nivel de la variable control que durante la fase A_1 . Este sería un primer uso de la variable control: servir como punto de comparación para evaluar los efectos residuales. El segundo uso de la variable control es el de reafirmar (estimar en una forma independiente, "cruzada") el efecto de la VI. Si la variable experimental fuese la única registrada, se podría plantear la objeción, poco factible, de que el cambio observado durante la fase B pudiera deberse a variables extrañas que actúan conjuntamente con la VI. Dicha objeción podría tener mayor credibilidad en situaciones de investigación aplicada, ya que, efectivamente, pudiese haber un número de variables fuera del control absoluto del experimentador que afectaran a la VD. Cuando se incluye la variable control, y ésta es independiente de la variable experimental, es posible contrarrestar la objeción. Si durante la fase B la variable control se mantuviera a su nivel de línea base mientras la variable experimental tuviera un cambio notable, la objeción podría descartarse con facilidad. (Véase Ribes, 1972, para más detalles.)

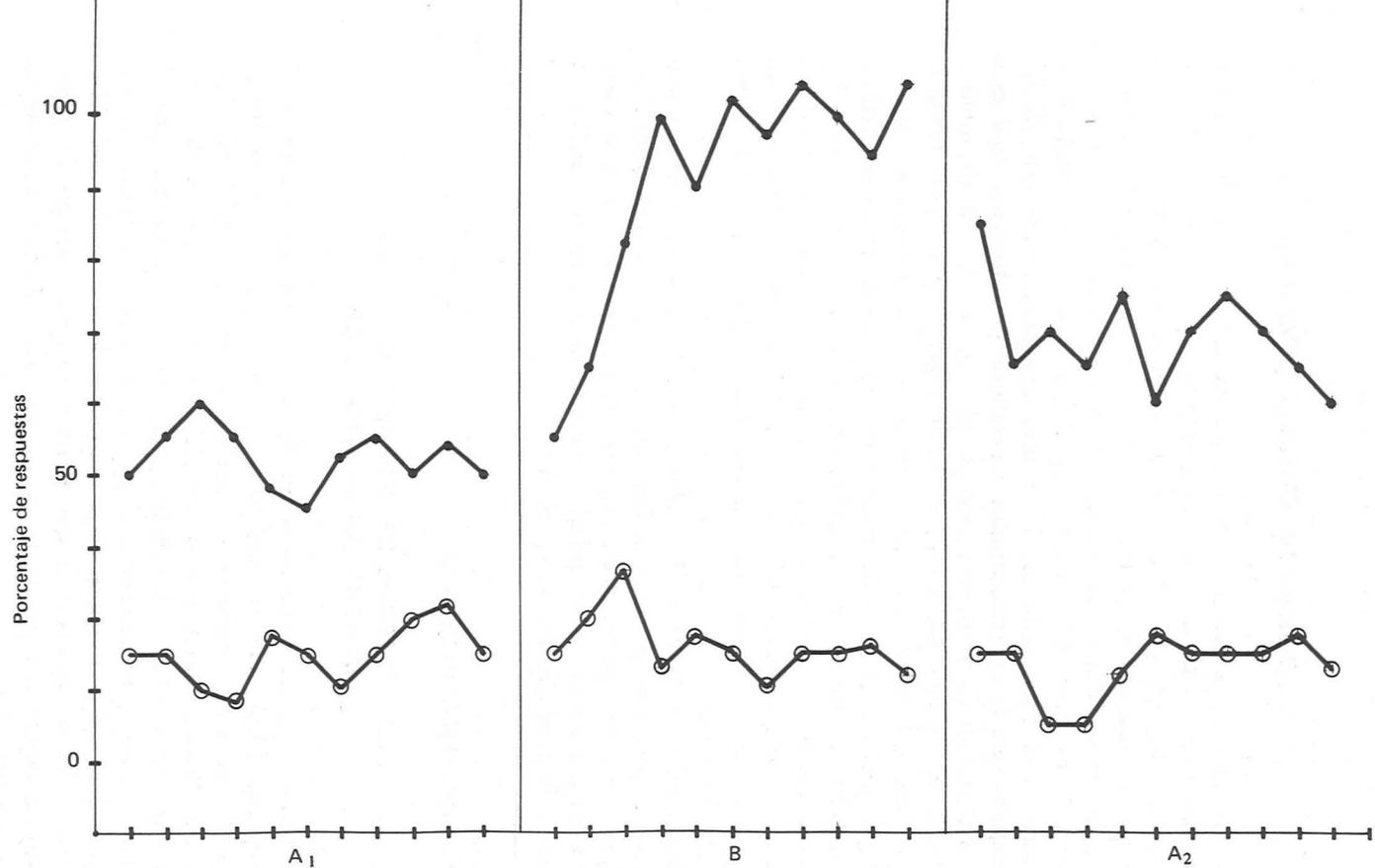


Figura 31. Datos ficticios que muestran un diseño A-B-A, en el cual se compara una variable experimental (círculos llenos), y una variable control (círculos en blanco).

9.2. DISEÑO DE CRITERIO CAMBIANTE

Este diseño fue propuesto por Hall y sus colaboradores (Hall, 1971) y está basado en el proceso de moldeamiento. En este diseño el investigador fija de antemano un criterio inicial a ser alcanzado por la conducta del sujeto. Una vez que dicho criterio se alcanza, mediante el uso de alguna técnica de control de la conducta se fija un nuevo criterio más exigente. El criterio se modifica conforme la conducta se estabiliza en el nivel anterior. Para usar apropiadamente este diseño, es necesario que el experimentador especifique de antemano cuál es el nivel terminal de la respuesta manipulada; o, en el caso de moldeamiento de una nueva conducta, se deben especificar las características de la conducta terminal, tales como su topografía, frecuencia, etc.

Weiss y Hall (1971) utilizaron con éxito este diseño para reducir gradualmente el número de cigarrillos fumados por un estudiante. Las consecuencias por fumar más cigarrillos que los fijados por el criterio en una fase determinada del diseño era la de tener que romper un billete de 1 dólar en el momento en que esto sucediera. El sujeto nunca tuvo que hacerlo.

Hall (1971) y Weiss y Hall (1971) consideraron que este estudio carecía de precisión por no incluir una reversión o una línea base múltiple. Por otra parte, arguyen que puede demostrarse que existe control experimental (causalidad) cuando la conducta cambia de acuerdo a lo especificado en el diseño.

II. CLASE TRADICIONAL

9.3. DISEÑOS DE MEDIDAS REPETIDAS DENTRO DEL MISMO SUJETO

Aun cuando en la literatura se encuentran referencias a diseños "de medidas repetidas", "dentro del sujeto", "un sujeto como su propio control" y cuestiones similares, en esta sección se les ha calificado bajo el título de "medidas repetidas dentro del mismo sujeto", a fin de evitar confusiones con el empleo previo que hemos hecho de los conceptos de "un sujeto como su propio control" y el de "medidas repetidas". Ambos conceptos han sido utilizados en el contexto de los diseños del AEC. Ambos se refieren al hecho de que el mismo sujeto o grupo de sujetos se miden u observan repetidamente a través de diferentes fases de un diseño experimental.

El concepto genérico de diseños reversibles es, generalmente, más preciso para describir los diseños del AEC que incluyen “un sujeto como su propio control” y toman “medidas repetidas”. Como el diseño A-B no es propiamente un diseño reversible, en ocasiones se usa cualquiera de estos conceptos para referirse a él. Es por eso que enfatizamos la terminología “dentro del mismo sujeto” para hacer referencia a los diseños tradicionales. Ahora, examinaremos las diferencias fundamentales entre los diseños tradicionales y los del AEC que usan “medidas repetidas”; veamos como se emplean las medidas repetidas en un diseño tradicional.

Quando se trata de reducir el monto de “error” (variabilidad) debido a diferencias individuales entre los sujetos que componen un grupo o condición, se recurre a hacer que el mismo sujeto sea sometido a todas las diferentes condiciones. Esta es una de las formas más comunes de diseño tradicional de medidas repetidas y sirve de base para diseños como el de cuadrados latinos. Johnson y Solso (1971) y Hayes (1963) señalaron que el principal uso de este tipo de diseño se lleva a cabo en experimentos de psicofísica, en los cuales el mismo sujeto debe ser sometido a diferentes niveles de estimulación.

Algunos investigadores, como Winer (1962) o Myers (1972), conceptualizan a un diseño con k tratamientos y n sujetos como un factorial de $k \times n$, donde k es el número de valores de una VI y n son los “valores” de la “VI sujetos”. De acuerdo con esta formulación, la interacción sujetos \times tratamientos produce la variabilidad (“término de error”) contra la cual se deben evaluar los efectos de la “VI tratamientos” y los de la “VI sujetos”. Como se puede ver, esta última es un seudofactor o seudovariante independiente, ya que no es manipulada de manera directa por el experimentador y sus “valores” están definidos vagamente.

Tanto el diseño de medidas repetidas dentro del mismo sujeto como sus derivados presentan, entonces, dos ventajas principales sobre otros diseños tradicionales: reducen la variabilidad dentro de las condiciones y disminuyen considerablemente el número de sujetos. Por otra parte, presentan el riesgo de producir efectos residuales (de práctica o fatiga, según Johnson y Solso) o de orden. También tienen la limitación de producir datos estadísticamente dependientes entre sí. Cuando se compara a este diseño con su contraparte del AEC, se puede notar la ausencia de dos características fundamentales en éstos; la línea base conductual y la reversión.

La ausencia de estas dos características excluye absolutamente la conceptualización de este diseño por sí solo como parte de los diseños del AEC.

Namboodiri (1972) estudió el problema de la dependencia estadística entre las medidas repetidas en los mismos sujetos y señaló algunos de los cambios en diseño que se han utilizado para analizar la extensión de dicha dependencia. Uno de los métodos empleados para disminuir los efectos residuales es el de incluir periodos de descanso entre los diferentes tratamientos, aun cuando no siempre es posible o deseable incluir dichos periodos. Otro método para estimar tanto los efectos directos como los residuales, es el de elaborar diseños balanceados en los cuales cada condición es precedida por todas las demás con igual frecuencia. La principal desventaja de dichos diseños [según Lucás (1957) y Cochran y Cox (1957) citados en Namboodiri (1972)] radica en que las estimaciones de los efectos residuales son menos precisas que las de los efectos directos. Un tercer método para reducir el problema de los efectos residuales es el de incluir un periodo extra al final del último tratamiento que recibe el sujeto en un diseño balanceado. El diseño resultante se conoce como *diseño de un periodo extra* y la tabla 3 muestra una versión de él, incluyendo cuatro sujetos y tres condiciones experimentales. Aunque la deficiencia encontrada en los diseños balanceados se reduce de manera considerable con la inclusión del periodo extra, no se la suprime completamente.

Tabla 3.

DISEÑO DE UN PERIODO EXTRA QUE INCLUYE TRES CONDICIONES
EXPERIMENTALES Y CUATRO SUJETOS

Periodos	Sujetos			
	1	2	3	4
1	A	B	C	A
2	B	C	A	C
3	C	A	B	B
4	C	A	B	B

NOTA: Las letras A, B y C denotan las condiciones experimentales.

Quando se supone que el simple paso del tiempo puede tener algún efecto sobre la VD, se analiza cuál es la tendencia de los datos en las medidas repetidas. A este efecto se le llama "envejecimiento" y resulta útil para el ejercicio de "encaje de curvas", que consiste en encontrar la función matemática que mejor describa una serie de datos.

El diseño de medidas repetidas dentro del mismo sujeto es un diseño básico para formar diseños más complejos dentro de la clase tradicional. Entre dichos diseños se encuentran los ya mencionados diseños balanceados y contrabalanceados, los diseños *entre-dentro*, los diseños jerárquicos, los cuadrados latinos y otros cuadrados.

9.4. DISEÑOS ENTRE-DENTRO

Este tipo de diseño incluye características de los diseños de dos o más grupos aleatorios y de los diseños de medidas repetidas. Myers (1972) considera a este tipo como uno de los más frecuentemente utilizados en psicología. La razón de tal difusión la relaciona a la incorporación de las características antes mencionadas: por un lado permite la comparación de diversos niveles de una VI (parte "entre" del diseño), y por el otro permite la evaluación de los cambios ocurridos en la conducta de un mismo sujeto (parte "dentro" del diseño). El número de VI que participan en un diseño determina su nombre; por ejemplo, un diseño "1 entre -1 dentro" contiene una variable de comparación entre varios grupos y una variable en la cual se mide repetidamente al mismo sujeto.

Se debe hacer notar que las medidas repetidas en el mismo sujeto pueden tomarse para evaluar el efecto del paso del tiempo en el comportamiento de los sujetos en las diferentes condiciones, o bien, para evaluar los efectos de una o más VI aplicadas a cada grupo de sujetos. También se puede variar el orden de los tratamientos que recibe cada grupo de sujetos en la forma que se hace en los diseños contrabalanceados. Es importante señalar que, aun cuando se mida repetidamente a un mismo sujeto, no se utiliza la metodología de la línea base; por tanto, la evaluación de los efectos de tratamientos sucesivos en los mismos sujetos va a tener los mismos problemas presentes en otros diseños de medidas repetidas, básicamente efectos residuales y efectos reactivos.

Cuando se mide repetidamente a los sujetos sin aplicar ninguna variable aparte de la VI que se maneja en la parte "entre" del diseño, el rol del tiempo como VI es cuestionable. Algunos investigadores podrán considerar a este diseño como *1 entre -1 dentro*; otros solamente lo considerarán como un diseño de n grupos aleatorios con medidas repetidas. Por otra parte, cuando se trata a los sujetos de acuerdo a las condiciones experimentales determinadas por la VI que hace comparaciones entre grupos y también se aplican diferentes valores de una segunda VI a los sujetos de cada grupo, se puede observar que este tipo de diseño tiene una estructura similar a la de un diseño factorial, en el cual se encuentran dos efectos principales y una interacción. La

diferencia principal es que, mientras en este último las observaciones dentro de cada celdilla (condición experimental) son independientes del resto de las celdillas (cada grupo de sujetos sólo recibe una combinación particular de valores), en un diseño entre-dentro solamente las celdillas correspondientes a cada grupo son independientes de las de los otros grupos, pero las celdillas dentro de cada grupo no lo son. La figura 32 puede servir para aclarar este punto.

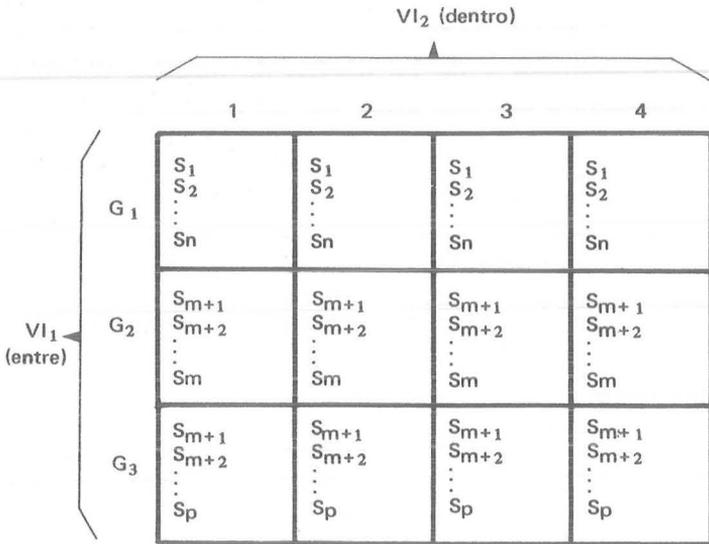


Figura 32. Esquema de un diseño 1 entre —1 dentro con tres valores de la VI_1 "entre" (tres grupos diferentes de sujetos) y cuatro valores de la VI_2 "dentro" (cada sujeto en cada uno de los tres grupos recibe los cuatro valores de la VI_2).

Este diseño es más útil para investigar problemas conductuales que aquellos diseños en los cuales solamente se pide una vez a cada sujeto, ya que permite "seguir" a un grupo de sujetos durante un tiempo determinado. Además, es más amplio que un diseño de medidas repetidas con un solo grupo de sujetos, ya que permite la comparación entre grupos tratados de diferentes maneras.

9.5. DISEÑOS JERÁRQUICOS (GRUPOS DENTRO DE TRATAMIENTOS)

Este tipo de diseño se usa frecuentemente en la investigación en el área educativa o, en general, cuando se trabaja en instituciones donde

los grupos de sujetos ya están formados (en forma de “clases”, “salas”, “compañías”, etc.). En un número de situaciones no es posible tratar a los sujetos independientemente del grupo al que pertenecen. Esto se puede deber a restricciones institucionales o a exigencias en la investigación aplicada, en las cuales se parta del supuesto de que los grupos pueden constituir unidades experimentales “naturales” (por ejemplo, “el grupo 2º B”, “la sala de esquizofrénicos” o “la 1ª compañía”). Es obvio que en ejemplos de este tipo los procedimientos de asignación de los sujetos a los grupos no depende en absoluto del investigador, y por tal razón el diseño se debería considerar como preexperimental.

Sin embargo, el diseño jerárquico no se limita necesariamente a este uso típico en la investigación aplicada. El diseño en general permite que parte de un tratamiento se “anide” a otro tratamiento. Esto significa que un diseño jerárquico podría utilizarse para la manipulación de dos o más VI. Entonces, su forma general sería la de un tratamiento dentro de otro tratamiento, y así sucesivamente, depen-

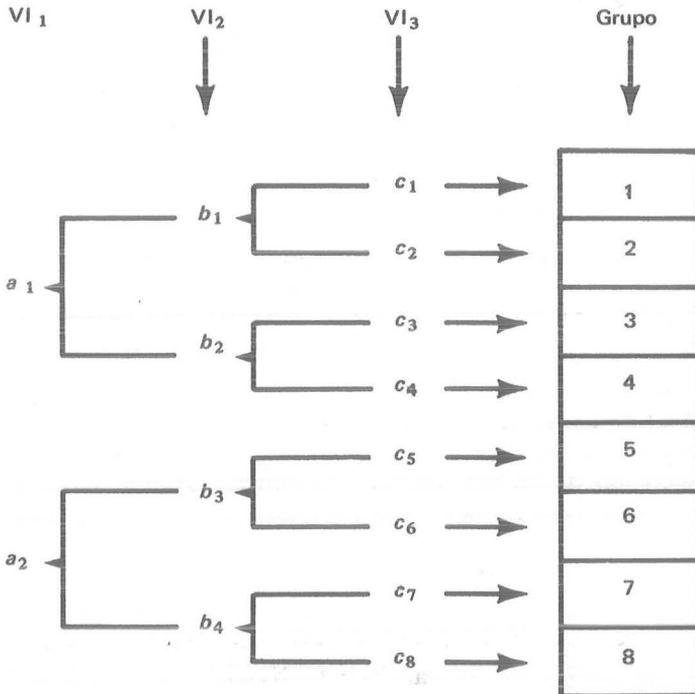


Figura 33. Esquema de un diseño jerárquico con tres VI: la primera con dos niveles, la segunda con cuatro niveles y anidada a la primera, y la tercera con ocho niveles y anidada a la segunda.

diendo del número de VI. La figura 33 presenta un diseño jerárquico con ocho niveles de la VI_3 anidados a cuatro niveles de la VI_2 , y éstos, a su vez, anidados a dos valores de la VI_1 . A diferencia de un diseño factorial equivalente con el mismo número de valores por VI que requeriría 64 grupos de sujetos, este diseño solamente necesitaría ocho.

Lindquist (1953), al analizar las aplicaciones del diseño a la educación, señaló dos usos principales del diseño: *a*) cuando es impracticable administrar diferentes tratamientos a distintos sujetos del mismo grupo y cada tratamiento se aplica a una serie independiente de grupos, y *b*) cuando cada grupo recibe todos los tratamientos de acuerdo a una replicación aleatoria (una forma de balanceo). El diseño, utilizado de cualquiera de estas dos formas, presenta más desventajas de las que admite Lindquist, entre ellas: *a*) su baja eficacia (señalada por Lindquist), sobre todo si la diferencia entre los grupos (presumiblemente antes del tratamiento) es demasiado grande; *b*) permitir que los efectos de la VI se confundan con los de VE, y *c*) permitir que los efectos sistemáticos de la selección produzcan diferencias indeseables en los grupos. Usado en la investigación preexperimental, el diseño presenta las ventajas de ser relativamente económico y de permitir una evaluación burda de los efectos de ciertos tratamientos.

9.6. DISEÑOS INCOMPLETOS

Hay ocasiones en las cuales el investigador está especialmente interesado en parte de la información que un diseño puede producir, pero no en toda la información obtenible. Cuando esto sucede, el uso de un diseño incompleto es preferible al de un diseño completo, la información de poco interés simplemente no se obtiene, ahorrándose así el investigador un esfuerzo innecesario. Hay otras situaciones en las cuales un diseño incompleto es el indicado: una de ellas es la de evitar la ejecución de un experimento sumamente grande y costoso, sacrificando ciertas condiciones experimentales en favor de otras. Otra situación tiene que ver con algunas suposiciones que hace el investigador de antemano, con tal de simplificar la naturaleza de su experimento. Dichas suposiciones pueden ser ciertas o no; en el primer caso, el investigador será afortunado en suponer que ciertos efectos resultarán nulos, y que realmente lo sean; en el segundo, se estará arriesgando a perder información que le podría resultar realmente útil, en cuyo caso puede estar distorsionando su investigación. A continuación consideraremos algunos casos específicos de diseños incompletos.

9.6.1. Diseños factoriales incompletos

Un término comúnmente usado para referirse a estos diseños es el de "diseños factoriales fraccionales". Se les denomina así porque solamente se incluye una fracción de las condiciones posibles. Ya Fisher (1960) señalaba la posibilidad de hacer replicaciones de un experimento, utilizando sucesivamente diferentes fracciones de un diseño factorial. Kirk (1972) señaló cuatro condiciones que se deben cumplir para usar un diseño factorial incompleto: *a*) cuando el número de VI es grande (por lo menos 4 o 5); *b*) cuando el experimento es prohibitivamente grande; *c*) el número de niveles por VI debe ser igual, y *d*) debe haber alguna razón a priori para descartar las interacciones más complejas. A esto se pueden añadir dos posibilidades más: *e*) algunas de las interacciones más complejas no tienen una interpretación experimental adecuada o son incomprensibles para el experimentador, y *f*) algunas de las combinaciones (celdillas) son de escaso interés para el investigador.

Al utilizar alguna fracción de un diseño factorial (como 1/2 o 1/3) generalmente se sacrifican las interacciones más complejas, pero se obtiene información fundamental acerca de los efectos principales y, en la mayoría de las veces, acerca de las interacciones más simples. Desde el punto de vista del costo de una investigación (tanto en términos del número de observaciones como en términos monetarios), este tipo de diseño puede ser sumamente útil, ya que puede proporcionar información básica en cantidades sustanciales, implicando un esfuerzo mucho menor que el que requeriría el diseño factorial completo. Cuando se compara esta reducción en costo con la información que podrían proporcionar las interacciones complejas, se advierte que esta última debería ser sumamente importante para preferir un diseño factorial completo (lo cual es improbable). Estas consideraciones son válidas, desde luego, si se condicionan al punto *a* de Kirk: es evidente que con diseños bifactoriales, o quizá trifactoriales, es preferible obtener toda la información. Uno de los usos que señala Kirk (1972) para este tipo de diseños es en el campo de la investigación exploratoria: se pueden investigar varias VI inicialmente y, en base a los resultados obtenidos, dirigir la investigación subsecuente. Entre las desventajas que encontró este mismo investigador están las dificultades en la interpretación de los resultados y la necesidad de suponer que ciertas interacciones son despreciables.

Según McLean (1967), el experimentador debe resolver los siguientes problemas para hacer un uso adecuado del diseño: decidir cuál

es el modelo que corresponde a su problema (por ejemplo, si una interacción no es importante o realista puede eliminarla del modelo y simplificar el diseño), buscar un balance entre el "monto de error" (posible) de los diferentes efectos y buscar ortogonalidad, es decir, que la estimación de los efectos de una variable no sea afectada por los de otras variables.

9.6.2. Cuadrados latinos

Un cuadrado latino es un diseño incompleto, que presenta las siguientes características: *a*) cada unidad experimental recibe todos los tratamientos experimentales (o combinaciones de sus valores); *b*) el número de unidades experimentales (frecuentemente un solo sujeto) es igual al número de VI(s); *c*) cada tratamiento (o combinación) le es dado al sujeto sólo una vez (cada tratamiento aparece solamente una vez en cada fila del diseño), y *d*) cada tratamiento (o combinación) se da únicamente una vez en cada ocasión (cada tratamiento aparece solamente una vez por cada columna del diseño).

La figura 34 muestra un diseño cuadrado latino derivado de un diseño factorial de $3 \times 3 \times 3$ (siguiendo los pasos de McNemar, 1969). En el diseño factorial (misma figura) se marcaron con una *X* aquellas celdillas que aparecen en el cuadrado latino.

Como se puede notar, el valor 1 (de las tres VI) aparece un total de nueve veces, al igual que los valores 2 y 3. Hay un balance respecto a las contribuciones de cada VI para la formación del cuadrado latino; asimismo, aunque hay un balance respecto a los efectos principales (Winer, 1962), no lo hay respecto a las interacciones. Se puede apreciar, también, que el diseño resultante cumple con las especificaciones de un cuadrado latino planteadas con anterioridad.

La formación de este cuadrado latino (o de cualquier otro cuadrado latino en general) implica que las celdillas de éste son una muestra representativa de las celdillas del factorial completo. Específicamente, al usar la combinación de valores 111 ($VI_1 = 1, VI_2 = 1$) y ($VI_3 = 1$) en la primera celdilla, se está implicando que esta combinación sustituye o representa (\leftarrow) adecuadamente a las combinaciones 111, 121 y 131 (tanto en dirección horizontal como en dirección vertical). A continuación (tabla 4) se presenta la lista completa de dicho muestreo.

El cuadrado latino es un tipo de diseño atractivo, ya que reduce el esfuerzo que habría que hacer para obtener la información de un diseño completo; sin embargo, deben tomarse precauciones ya que a medida que aumenta el tamaño del diseño, es menos probable la im-

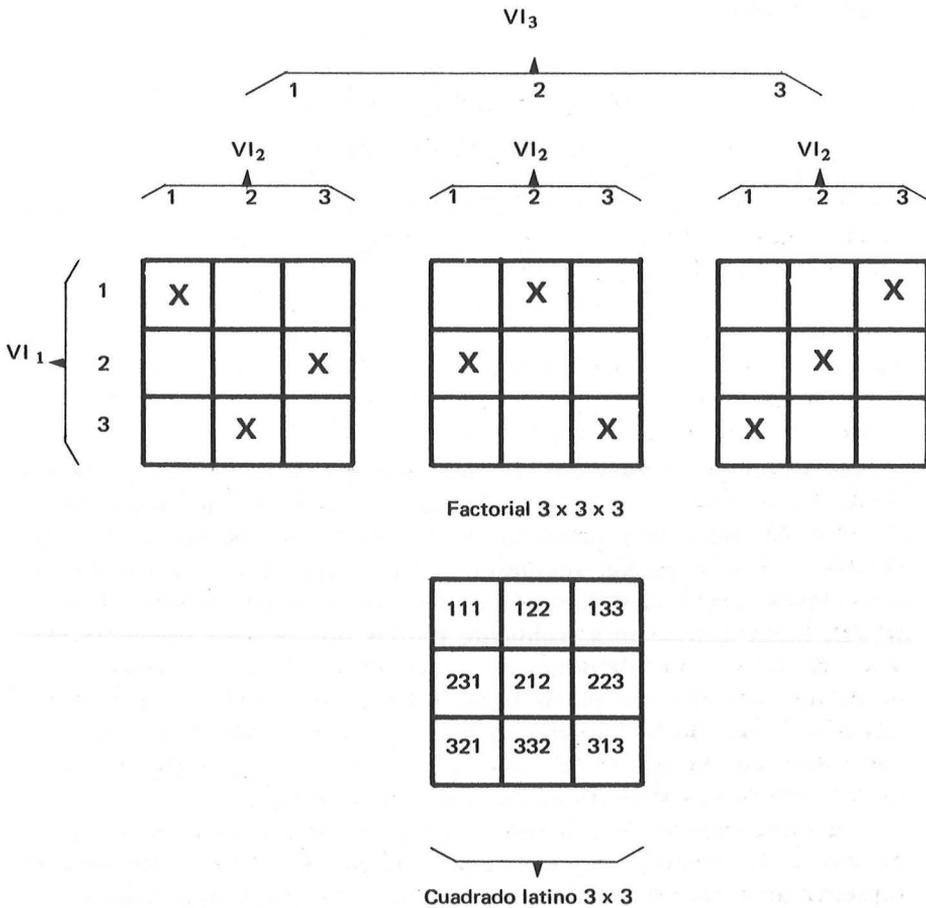


Figura 34. Celdillas de un diseño factorial de $3 \times 3 \times 3$ (arriba) que se utilizaron para formar un cuadrado latino de 3×3 . Los números en las celdillas del cuadrado latino indican las combinaciones de los valores de las tres VI.

Tabla 4.

LISTA DE LAS CELDILLAS DE UN DISEÑO CUADRADO LATINO DE 3×3
 QUE "REPRESENTAN" UN DISEÑO FACTORIAL DE $3 \times 3 \times 3$

111, 121, 131	←— 111
112, 122, 132	←— 122
113, 123, 133	←— 133
211, 221, 231	←— 231
212, 222, 232	←— 212
213, 223, 233	←— 223
311, 321, 331	←— 321
312, 322, 332	←— 332
313, 323, 333	←— 313.

plicación de representatividad de un diseño completo por medio de un diseño incompleto. Cuando las celdillas seleccionadas para aparecer en el cuadrado latino no representan a las celdillas de un diseño factorial completo, la información resultante puede variar desde ligeras desviaciones en la magnitud o dirección de los efectos de una o más VI sobre la VD, hasta desviaciones realmente burdas que pueden presentar versiones completamente deformes de las relaciones entre variables.

Al respecto, el lector puede hacer el ejercicio siguiente: construir un esquema de un diseño factorial de $3 \times 3 \times 3$ como el de la figura 34, e inventar valores para cada una de las celdillas. Estos valores representan las medias aritméticas de cada grupo de sujetos en el diseño. Luego, puede obtener medias de las filas 1, 2 y 3 y trazar una gráfica con ellas. (Esta gráfica representaría el efecto de la VI_1 .) En seguida puede obtener las medias de las columnas, y al graficarlas tendrá una representación del efecto de la VI_2 sobre la VD. Finalmente, puede obtener las medias de cada uno de los cuadros para ver el efecto de la VI_3 .

El siguiente paso es el de graficar solamente los valores de las celdillas del cuadrado latino (se pueden usar las mismas de la figura 34, y para simplificar el problema se puede suponer que son las medias de grupos de sujetos). Al hacer esto, podrá construir tres gráficas, que correspondan a las filas, las columnas y los cuadros. Hay tres cosas que pueden ocurrir, en general: a) obtener un acuerdo general entre

estas gráficas y las obtenidas con el diseño completo; *b*) lograr un acuerdo parcial (una o dos de las gráficas son similares a las del diseño factorial), o *c*) obtener un completo desacuerdo. Si ahora forma un segundo cuadrado latino (por ejemplo, intercambiando dos filas o columnas del cuadrado anterior), obtendrá un resultado diferente, a menos, claro, que todos los números de las celdillas fueran iguales. Este nuevo cuadrado puede representar adecuadamente al diseño completo, pero también pudiera no representarlo en absoluto, o representarlo sólo parcialmente. Se pueden seguir construyendo diferentes cuadrados latinos con resultados variables. Cuando las diferencias entre las celdillas del diseño completo son grandes, hay el riesgo de que las celdillas del cuadrado latino den una representación verdaderamente deforme de aquéllas. La dirección y, sobre todo, la magnitud de los efectos, pueden estar total o parcialmente distorsionados. Cuando no hay verdaderos efectos de las VI en el diseño completo, el riesgo de representación inadecuada disminuye; entonces, el cuadrado latino indicará que no hay efectos de las VI. Esto, sin embargo, plantea una nueva incógnita: cuando un cuadrado latino indica que no hay efectos importantes de las VI, esto se puede deber a que efectivamente no los haya (en cuyo caso no hay problema), pero también se puede deber a que, habiendo efectos reales, la combinación particular de celdillas del cuadrado latino sea insensible a dichos efectos. ¿Cuál de estas alternativas es la correcta? Una manera de averiguarlo es, como ya se sugirió, la de desarrollar diferentes cuadrados latinos; si sus resultados son inconsistentes entre sí, esto complicaría aún más las cosas, e incluso, en caso de que sean consistentes, no hay una garantía de que esta combinación particular de cuadrados latinos no sea una improbable combinación de los únicos resultados consistentes.

Lo que aquí se ha planteado como un mero ejercicio, algunos investigadores realmente lo han llevado a cabo como una forma de replicación, a fin de asegurarse que tienen datos confiables (véanse Winer, 1962, o a Edwards, 1968, para una explicación al respecto). El mismo Winer apuntó que este diseño se usa en la investigación biológica, con el fin de controlar las diferencias individuales entre las unidades experimentales, y que en las ciencias de la conducta se le emplea, además de la forma ya indicada, para "controlar" los efectos del orden y como el bloque básico para la construcción de diseños más complejos. Al igual que en el diseño anterior, uno de los principales problemas con el cuadrado latino es el de tener que suponer que ciertas interacciones son inexistentes o despreciables. McNemar (1969) indicó que para poder evaluar un cuadrado latino es necesario suponer que la interacción entre orden y tratamiento es igual a cero. Asimismo, señaló que hay una división de

opiniones respecto al valor del cuadrado latino en la investigación psicológica, siendo su propia opinión en el sentido de que la sustitución de un factorial completo por medio de un cuadrado latino presenta desventajas que sobrepasan desproporcionadamente la única ventaja de éstos: permitir el uso de números pequeños de sujetos.

Otro cuadrado relacionado con el cuadrado latino es el greco-latino. Este diseño se puede considerar como una extensión del cuadrado latino, ya que se forma sobreponiendo dos cuadrados latinos, manteniendo las características de aquél y añadiendo la restricción de que cada letra griega (una nueva VI) debe aparecer en combinación con cada letra latina una sola vez. Winer (1962) consideró que un cuadrado greco-latino es realmente una replicación fraccional de un factorial con cuatro VI (fila, columna, letra latina y letra griega) y que su utilidad está limitada a aquellas situaciones en las cuales las interacciones son despreciables. Finalmente, a veces aparecen en la literatura referencias al "cuadrado Youden". Este es un caso particular de los diseños balanceados incompletos. Este diseño se construye con bloques (sujetos "homogeneizados") y tiene la característica de que el número de bloques que aparece en el diseño siempre es igual al número de tratamientos.

9.7. DISEÑOS DE BLOQUES ALEATORIOS

Estos diseños tuvieron su origen en la agricultura, con el siguiente planteamiento: un bloque era un cuadrado de tierra dividido en varias parcelas. El principal problema para el experimentador consistía en controlar el grado de semejanza entre las parcelas y entre los bloques (la parcela 1 es más semejante a la 2 que a la 3, y así sucesivamente) como una variable cuyo efecto se pudiese confundir con el de la VI. La suposición básica, generalmente confirmada en problemas agrícolas, es la de que hay mayor homogeneidad dentro de los bloques que entre ellos, facilitándose así la evaluación de los efectos de los tratamientos.

El problema de la semejanza entre parcelas se controlaba por medio de la aleatorización. Los diversos niveles de la VI se asignaban al azar a las parcelas dentro de cada bloque. Se realizaba el experimento (por ejemplo, usando distintas dosis de un fertilizante) y se analizaban las diferencias entre los niveles de la VI y las diferencias entre bloques, las cuales indiscutiblemente tenían sentido en esta clase de investigación.

En el campo de la investigación psicológica, por otra parte, se emplean estos diseños con el exclusivo propósito de reducir la variabilidad dentro de los bloques (grupos homogéneos de sujetos), ya sea de la

propia VD o de una variable relacionada a ésta. Desde luego, las bases sobre las cuales se divide a los bloques en “parcelas” no están definidas en absoluto. En algunos casos se intenta producir la homogeneidad de los bloques agrupando a los sujetos por “camadas”, y en otras ocasiones por apareamiento indirecto o correlacional. La suposición de que hay mayor homogeneidad dentro de los bloques que entre ellos, se puede cumplir en la medida en que dichos procedimientos sean efectivos, y, por supuesto, existe el riesgo de que no se cumpla en absoluto. Cuando éste es el caso, todo el procedimiento de agrupar a los sujetos por bloques resulta infructuoso.

Tal como en la investigación agrícola, se emplea el procedimiento de aleatorización para asignar los sujetos a los niveles de la VI. Si un experimentador enfatiza la importancia de la reducción de la variabilidad por medios esencialmente estadísticos (como sería el agrupamiento por bloques), entonces este tipo de diseño parecerá preferible a los de grupos aleatorios. Si, por lo contrario, el investigador prefiere reducir la variabilidad por medio del control experimental, lo más probable es que considere a este tipo de diseño como de escasa utilidad o artificioso. Una notable excepción sería el caso en que su investigación exploratoria haya demostrado que ciertos parámetros de respuesta de los sujetos se pueden agrupar “naturalmente” en varios niveles (una forma de selección de una VI) y que su estudio en la interacción con los niveles de la VI fuese interesante.

10

evaluación y comparación de los diseños experimentales

EN ESTE CAPÍTULO se hará una evaluación y comparación de las dos clases principales del diseño experimental en la investigación conductual: los diseños tradicionales y los diseños conductuales. El capítulo se divide en tres secciones: una, donde se evalúa a los diseños tradicionales, independientemente de los diseños del AEC; otra, donde se evalúa a los diseños del AEC, independientemente de los diseños tradicionales; y una tercera, donde se hace una comparación lo más exhaustiva posible entre ambas clases de diseños.

10.1. EVALUACIÓN DE LOS DISEÑOS TRADICIONALES

Para hacer esta evaluación es necesario considerar las características de los diseños tradicionales, comparando diversos diseños tradicionales entre sí. También es necesario considerar la validez de la aproximación metodológica que fundamenta a este tipo de diseños. Finalmente, hay que considerar a los diseños alternativos (en este caso, los diseños conductuales). Las dos primeras consideraciones se estudian en esta sección y la última se analiza en la sección 10.3.

10.1.1. Características de los diseños tradicionales

Dentro de las características de los diseños a considerar tenemos: *a)* su *informatividad* (cuánta información proporcionan); *b)* su “*costo*” (en términos del número de observaciones); *c)* su poder; *d)* sus *restricciones metodológicas*, tales como el mínimo de sujetos por condición experimental o grupo, el mínimo de condiciones experimentales, etc.; *e)* sus *restricciones prácticas*, *f)* su propósito, y *g)* el *efecto de la estadística* sobre la forma del diseño o sus restricciones metodológicas. A continuación se analizará cada una de dichas características.

10.1.1.1. Capacidad informativa

Qué tan informativo es un diseño depende básicamente de qué tan complejo es (cuántas condiciones experimentales contiene). Además, también depende del número de VD bajo estudio y de la cantidad de observaciones (medidas repetidas) por unidad experimental (grupo).

Obviamente, los diseños más simples son los menos informativos (por ejemplo, el de dos grupos aleatorios o el de dos grupos apareados). Aunque los diseños más informativos se encuentran entre los más complejos, la complejidad no depende solamente del número de grupos contenidos en el diseño, sino, además, del número de VI y de la forma en que cada condición experimental se relaciona con las demás. Por ejemplo, un diseño de cuatro grupos aleatorios es, generalmente, más informativo que uno de tres grupos aleatorios; sin embargo, un diseño factorial de 2×2 es más complejo e informativo que uno de cuatro grupos aleatorios, a pesar de que ambos contienen el mismo número de condiciones. La primera diferencia radica en que el diseño factorial incluye más VI (2 contra 1). La segunda consiste en que sus condiciones experimentales están organizadas de tal manera que puede contestar más preguntas de investigación o poner a prueba más hipótesis (3 contra 1).

La complejidad también depende del número de mediciones u observaciones por unidad experimental; por ejemplo, compárense dos diseños factoriales con cuatro grupos cada uno. El primero es un 2×2 y el segundo es un $2 \times 2 \times 10$, donde el 10 significa que se tomaron diez mediciones (mediciones repetidas) por cada unidad experimental o grupo. Obviamente, el segundo diseño es más complejo, ya que produce más datos por unidad experimental y, por ende, es más informativo.

Por último, incluso un diseño experimental simple puede ser bastante informativo si se estudian simultáneamente varias VD empleando las mismas unidades experimentales (grupos). Desde luego, un diseño experimental complejo en el cual se midan simultáneamente varias VD sería todavía más informativo.

10.1.1.2. "Costo"

Por "costo" de un diseño se entiende en esta obra: a) el número de observaciones por unidad experimental; b) el número de sujetos por unidad experimental; c) el número de condiciones experimentales, y d) el número de VD. El número total de observaciones por diseño se obtiene multiplicando $a \times b \times c \times d$ y esto nos da el costo exacto de un determinado diseño. Por tanto, el costo de un diseño se puede definir como el número total de observaciones por diseño.

En la tabla 5 se compara el costo de varios diseños tradicionales.

Tabla 5

Costo (número de observaciones por diseño) de cuatro diferentes diseños tradicionales.

Índice	DISEÑO			
	C-E	C-E ₁ -E ₂	C-E × 10	3 × 3 × 3
Número de observaciones por unidad experimental	1	1	10	1
Número de sujetos por unidad experimental	20	20	20	20
Número de condiciones experimentales	2	3	2	27
Número de VD	1	1	1	1
Costo	40	60	400	540

NOTA: C-E = Grupo control — grupo experimental,
 C-E₁-E₂ = Grupo control-grupo experimental 1-grupo experimental 2,
 C-E × 10 = Grupo control-grupo experimental con diez mediciones repetidas, y
 3 × 3 × 3 = Diseño factorial de 3 × 3 × 3.

Este concepto de costo se relaciona, entonces, al esfuerzo que hay que invertir para obtener un número de observaciones por diseño. No se intenta darle un significado monetario, aun cuando sería relativamente fácil hacer una conversión, determinando cuánto dinero se debe invertir en cada experimento para obtener cada observación.

10.1.1.3. Poder

El poder de un diseño se refiere a la relación que guarda el *costo* con su *capacidad informativa*; un diseño será más poderoso cuando por el mismo costo genera más información. Una comparación típica sería la comparación de un diseño factorial de 2×2 con el uso repetido (dos veces) de un diseño de dos grupos aleatorios. El primero sería más poderoso puesto que por el mismo costo pondría a prueba tres hipótesis contra sólo dos de la segunda aproximación.

Otra comparación relevante sería entre un diseño de dos grupos aleatorios con diez veces más sujetos y un diseño de dos grupos aleatorios con diez veces más mediciones repetidas (y diez veces menos sujetos). Como se ve en la tabla 6, el costo de ambos es el mismo y, sin embargo, el segundo es más informativo, puesto que además de poner a prueba la hipótesis respecto al efecto del tratamiento (la cual también pone a prueba el otro diseño), pone a prueba una hipótesis (o contesta una pregunta) respecto a la tendencia de los datos en ocasiones sucesivas.

Tabla 6

Comparación de costos entre un diseño de dos grupos aleatorios y un diseño de dos grupos aleatorios con mediciones repetidas.

Índice	DISEÑO	
	C-E	C-E \times 10
Número de observaciones por unidad experimental	1	10
Número de sujetos por unidad experimental	100	10
Número de condiciones experimentales	2	2
Número de VD	1	1
Costo	200	200

Es conveniente señalar una objeción válida a la comparación anterior: el diseño con menor número de sujetos por unidad experimental contendrá una mayor variabilidad de muestreo, y los datos obtenidos con él contendrán una mayor proporción de error de medida en lo que se refiere a indicar los "verdaderos" parámetros o valores representativos de cada grupo comparado. Esto, en suma, los haría menos confiables; por tanto, el segundo diseño sería más poderoso, pero, bajo estas circunstancias particulares, también menos confiable.

10.1.1.4. Restricciones metodológicas

Éstas son las restricciones en las características de los diseños, su uso o evaluación, las cuales son producto de la aproximación metodológica asociada con este tipo de diseños. Aquí se considerarán las siguientes: *a)* el costo mínimo; *b)* el uso de grupos; *c)* el uso de la estadística; *d)* las escalas de medición, y *e)* la confiabilidad. Considerémoslas en ese orden:

a) Costo mínimo. Los diseños experimentales tradicionales deben tener un costo mínimo (es decir, el mínimo de observaciones que se deben hacer para obtener información útil). Éste se descompone de la siguiente manera:

- i)* El número mínimo de observaciones por unidad experimental es igual a 1. Cuando éste es el caso, se sobreentiende que dicha observación o medición se hace después de la aplicación de los tratamientos experimentales.
- ii)* El número mínimo de sujetos por unidad experimental. Este número no es fácil de determinar y dependerá grandemente de la técnica estadística con que se analicen los datos. En general, comoquiera que sea, números menores de diez serán rechazados por la mayoría de los experimentadores. (El mínimo comúnmente aceptado será entre 10 y 20.) En casos realmente excepcionales se podrían aceptar sólo cinco sujetos por grupo (por ejemplo, en aquellos casos en los que sea prácticamente imposible conseguir más sujetos, o en casos en los que se haya hecho un muestreo adecuado y los datos no muestren una variabilidad extremadamente grande o pequeña). Kirk (1968) afirmó que se deben considerar cinco factores, a fin de determinar el número de sujetos requeridos para poner a

prueba una hipótesis estadística: *a*) los efectos mínimos del tratamiento experimental a ser advertidos; *b*) el número de niveles de tratamiento; *c*) la varianza del error de la población; *d*) la probabilidad de cometer un error de tipo I,* y *e*) la probabilidad de cometer un error de tipo II.**

iii) El mínimo de condiciones experimentales. Este número es, por supuesto, igual a 2. Recuérdese que no se puede tener un diseño experimental más simple.

iv) El número mínimo de VD es igual a 1.

Por tanto, el costo mínimo comúnmente aceptado será entre 20 y 40 y el mínimo excepcional será de 10.

b) El uso de grupos. Este es el elemento básico de los diseños tradicionales; por esta razón, los grupos son considerados como *unidades experimentales*, y no los individuos que forman los grupos. Los datos que producen los grupos (sometidos a diferentes tratamientos) son fundamentales para la evaluación de los tratamientos y para decidir acerca de las hipótesis. La posibilidad de analizar los datos de un sujeto individual no se considera importante.

c) El uso de la estadística. Prácticamente todos los diseños experimentales tradicionales usan alguna técnica estadística para el análisis de los datos que con ellos se obtienen. Este uso puede ser de un nivel simplemente descriptivo o puede ser de un nivel inferencial. Generalmente se ponen a prueba hipótesis y rara vez se estiman parámetros (Simon, 1969). Los análisis gráficos o de otra clase son considerados como complementarios.

d) Escalas de medición. Probablemente a partir de la publicación del libro de Siegel (1970) en 1956 *** ha aumentado notablemente el énfasis en el uso adecuado de las escalas de medición para la obtención de datos en psicología. De acuerdo a las consideraciones de Siegel, solamente las mediciones que obedecen a una escala de intervalos o a una superior (de proporción) podrían calificar para ser utilizados en los diseños experimentales tradicionales. Un número cada vez mayor de investigadores tiene en cuenta esta restricción al planear y ejecutar sus experimentos. A pesar de esto, hay estadísticos como Lord (1972),

* Un error de tipo I es aquel en el cual se rechaza la hipótesis nula, cuando en realidad es cierta.

** Un error de tipo II es aquel en el cual no se rechaza la hipótesis nula, cuando en realidad es falsa.

*** Según Kirk (1972), esta tendencia comenzó diez años antes con la publicación de un trabajo de S.S. Stevens.

Gaito (1972), Boneau (1972) y Anderson (1972) que definitivamente rechazan dicha restricción y consideran que su naturaleza no es estadística, sino de medición.

e) *Confiabilidad*. Esta medición se puede conceptualizar como la consistencia de las mediciones a través del tiempo o mediante formas equivalentes (para Nunally, 1967, es la extensión en que las mediciones son repetibles). La confiabilidad se reporta con poca frecuencia en los estudios donde se usa algún diseño experimental tradicional. Cuando se utiliza una prueba psicométrica ya comercializada, simplemente se acepta la confiabilidad reportada por los editores de la prueba. Cuando la prueba es nueva, un investigador cuidadoso computará su confiabilidad y no la usará si no es aceptable; sin embargo, en pocas ocasiones aparece dicho dato en las revistas especializadas.

Cuando hay equipo automático de registro no hay problema de confiabilidad. Como la confiabilidad tradicional es, básicamente, un parámetro de naturaleza estadística (una correlación), no resulta típico en esta clase de diseños que se reporte la confiabilidad interobservadores en términos de la proporción de acuerdos, como se suele hacer en los diseños conductuales.

10.1.1.5. Restricciones prácticas

No es raro que un experimentador planee llevar a cabo un experimento o una serie de experimentos usando determinado diseño experimental y termine haciéndolo con un diseño mucho más simple. ¿Por qué? Las principales causas probablemente tendrán que ver con restricciones prácticas imprevistas. Estas restricciones pueden limitar severamente las opciones de un investigador y también pudieran determinar la selección de un diseño de una clase sobre un diseño de la otra. Es posible clasificar a dichas restricciones prácticas en tres categorías, a saber: a) sujetos, b) número de observadores o experimentadores, y c) presupuesto.

a) *Sujetos*. Hay dos clases de restricciones prácticas relacionadas con los sujetos:

- i) Número de sujetos. Este número requerido aumenta conforme crece la cantidad de condiciones en el diseño. No siempre se puede obtener el número de sujetos indicado por un diseño (por ejemplo, si se trabaja con voluntarios o en una institución pequeña) y esto obligaría a la reducción del número de condiciones experimentales. También se debe considerar el

número de sujetos por condición experimental, e independientemente de las consideraciones de costo mínimo, se sabe que los resultados serán más consistentes (menos variables y menos afectados por puntuaciones caprichosamente extremas) cuando hay un número grande de sujetos por condición experimental.

- ii) Disponibilidad temporal de los sujetos humanos. Obviamente, cuando se realizan experimentos con animales, esta restricción es mínima; sin embargo, con los humanos es importante considerar la cantidad de tiempo que los sujetos están disponibles. El hecho de que la gran mayoría de los experimentos con seres humanos sean con poblaciones de estudiantes impone ciertas limitaciones al tiempo que se puede contar con ellos; por ejemplo, un diseño con mediciones repetidas pudiera requerir 12-18 semanas (lo que generalmente dura un semestre académico) para tomarlas y pocos sujetos estarían dispuestos a participar tanto tiempo en un experimento. Incluso, pocos administradores o maestros (quizá con la excepción de aquellos en psicología) estarían dispuestos a ceder el tiempo de sus estudiantes por más de una ocasión; ésta es una restricción que hace preferible a un diseño de comparación entre grupos sobre uno con mediciones repetidas en el mismo sujeto. Para finalizar, se debe señalar que el número de sesiones varía dependiendo del diseño, pero, en general, pocas veces excede de quince y frecuentemente sólo requiere una o dos. La duración de la sesión experimental varía con el experimento, pero, como regla empírica, a medida que aumenta su duración, disminuye el número de sujetos dispuestos a participar, por lo cual se podría presentar un problema de selección. Dejando a un lado los diseños de mediciones repetidas, se puede decir que los diseños tradicionales requieren números moderados de sesiones experimentales.

b) *Número de observadores o experimentadores.* Conforme aumenta la complejidad del diseño, el número de sujetos o la cantidad de VD, aumentan los requerimientos de observadores o experimentadores calificados. Aun cuando no haya limitaciones monetarias, la falta de observadores o experimentadores calificados (generalmente de estos últimos, cuando se trata de diseños tradicionales) puede restringir considerablemente las posibilidades de una investigación y alterar la selección del diseño.

c) *Presupuesto.* La cantidad de dinero asignada a un experimento puede ser un factor determinante en la selección del diseño. La mayo-

ría de las veces un presupuesto pequeño sólo alcanzará para implementar los diseños más simples. Un diseño complejo puede demandar un presupuesto grande, puesto que implicará un mayor número de sujetos (los cuales reciben dinero por su participación en múltiples experimentos), un mayor número de experimentadores u observadores, una mayor cantidad de materiales, más aparatos de estimulación o medición, etc.

Un diseño factorial complejo, un diseño de grupos aleatorios con muchos grupos o un diseño con muchas mediciones repetidas implicará un fuerte presupuesto, fuera del alcance del investigador independiente con limitados recursos económicos. En consecuencia, todo investigador precavido contempla seriamente esta restricción práctica, antes de escoger un diseño de entre varios factibles.

10.1.1.6. Propósito

Esta característica se halla íntimamente relacionada con la estrategia de investigación adoptada por un experimentador, ya que uno de los propósitos para los cuales se emplea un diseño puede ser el de poner a prueba una hipótesis, o puede ser el de contestar una pregunta de investigación. Típicamente, los diseños tradicionales se usan con el primer propósito.

Otro propósito para hacer experimentos que pudieran tener un efecto sobre la selección de un diseño experimental es el de la estimación de parámetros. Quizá con la excepción de los investigadores en las áreas de modelos matemáticos, aprendizaje, procesamiento humano de información, o sensorpercepción, la mayoría de los estudios que emplean diseños tradicionales no tienen el propósito de estimar parámetros. Simon (1969) lamentó este estado de cosas en el campo de la psicología e indicó que no sólo se hace mucho más énfasis en la prueba de hipótesis que en la estimación de parámetros, sino que hay la tendencia a ignorar los parámetros en los cuales se basa una diferencia, indicando solamente si ésta es "significativa" o no. En el raro caso en que se lleve a cabo un experimento con el fin de estimar parámetros, el tipo de parámetro en cuestión puede afectar la selección del diseño (por ejemplo, si se intenta determinar un umbral perceptivo, sería preferible tener un número relativamente grande de grupos y no solamente dos).

Finalmente, el principal propósito de la experimentación conductual en su estado actual de desarrollo es el del establecimiento de relaciones funcionales inequívocas entre diferentes VD y VI, y esto implica tanto a los diseños tradicionales como a los conductuales.

Por lo que respecta al comportamiento de grupos de sujetos, los diseños experimentales tradicionales son capaces de llevar a dicho establecimiento de relaciones funcionales; sin embargo, como ya lo argumentó Sidman (1960), cuando se derivan predicciones de conducta individual a partir de datos grupales o cuando se trata de describir la conducta de un organismo particular en términos de una relación funcional de grupo, sólo se obtienen indicaciones demasiado vagas de la relación real, y algunas veces dichas indicaciones podrían ser totalmente erróneas (pocos sujetos se comportan como "el promedio" de un grupo; y aún menos sujetos se comportarán como los promedios de varios grupos).

10.1.1.7. Efecto de la estadística

La relación entre la estadística y el diseño experimental tradicional va más allá del mero análisis de los datos.

Se puede decir que el uso de un diseño experimental tradicional específico está condicionado por las suposiciones que sustentan a determinado modelo estadístico. Una de las suposiciones que sustentan a la mayoría de los modelos estadísticos es que las observaciones son estadísticamente independientes. Para poder asegurar que una diferencia entre dos medias aritméticas de dos grupos sea verdadera, se espera que ningún sujeto que participa en un grupo lo haga en el otro, a fin de lograr la independencia estadística. Por tanto, los diseños de grupos aleatorios deben en parte su estructura a esta suposición estadística. Otro ejemplo de la suposición de independencia estadística sobre los diseños tradicionales es el siguiente: cuando se mide repetidamente a un mismo sujeto o a un grupo de sujetos no se admite la posibilidad de que una medición sea independiente de la que la precedió o de la que le sigue. Si, a sabiendas de que hay cierta dependencia estadística entre las observaciones, se analizan los datos, se sabe que no se están obteniendo datos "limpios" y que será muy difícil precisar hasta dónde llegan tanto los efectos de un tratamiento como los efectos residuales de otro tratamiento o medición. Por esta razón, el investigador que haga estas consideraciones estadísticas tenderá a preferir diseños que se fundamenten en observaciones independientes, como son los diseños que se basan en la comparación de grupos de sujetos. Un tercer ejemplo del efecto de la estadística sobre los procedimientos del diseño experimental tradicional es proporcionado por aquellos diseños en que, con el fin de poder medir repetidamente a un mismo sujeto y aun así lograr algún grado de independencia estadística, incluye un "periodo de descanso" entre una medición y otra.

Una forma más indirecta en que la estadística puede tener un efecto sobre el uso de un diseño experimental es la siguiente: cuando se intenta emplear un diseño (por ejemplo, con un número relativamente grande de grupos) pero se considera que los datos se analizarán con el AVAR, el número de grupos se pudiera tener que reducir. Para usar este análisis se debe cumplir con los requisitos de continuidad de la VD e igualdad de las varianzas. A su vez, para aproximar lo más posible dichas suposiciones o requisitos, se debería tener números bastante grandes de sujetos en cada condición experimental; y si el número de sujetos disponibles es limitado, se tiene que reducir el tamaño del diseño. En tal caso, tenemos un ejemplo de un efecto indirecto de la estadística sobre la selección de un diseño a través de la producción de restricciones prácticas.

Ciertos diseños como los cuadrados latinos se construyeron en base a suposiciones como la de la ausencia de interacciones o del efecto nulo del orden de los tratamientos. Además, prácticamente ninguna técnica estadística se ocuparía de datos de grupos menores que cinco. Por otra parte, el proceso de la inferencia estadística depende de que las suposiciones en las que se basa un análisis no sean violadas; si ocurre una violación, esto podría dar lugar a una inferencia errónea, y toda la secuencia lógica que llevaría al rechazo de una hipótesis nula o de su alternativa se trastornaría gravemente.

10.1.2. Fundamento metodológico de los diseños tradicionales

En una forma simplificada y lo más general posible, se procederá a revisar el fundamento del diseño experimental tradicional, las operaciones o procedimientos que se producen como consecuencia de este fundamento y la forma en que éstas se encadenan lógicamente constituyendo un proceso de inferencia estadística-científica.

El fundamento principal del diseño tradicional radica en la equivalencia inicial entre las unidades experimentales. Con el fin de apreciar adecuadamente el efecto de un tratamiento experimental, es indispensable que los grupos sean lo más similares en todas las variables relevantes, y particularmente en la VD, antes de la inclusión del tratamiento.

Existen varios fundamentos de tipo secundario implícitos en este fundamento básico: primero, es posible formar unidades experimentales agrupando diferentes sujetos individuales; segundo, si éstos son tratados de modo uniforme, se comportarán en una forma reconociblemente uniforme; tercero, el o los momentos en que se les mide son típicos o representativos de otros momentos arbitrarios.

Las operaciones más comunes en el diseño tradicional para lograr este propósito son: la *aleatorización*, el *apareamiento*, las *técnicas de muestreo* y la aplicación de un *pretest*.

Ninguna de estas operaciones garantiza plenamente la equivalencia de los grupos implicados; incluso, en el remoto caso de que la equivalencia se logre, esto no significa que durante el transcurso del experimento (o aun inmediatamente antes o inmediatamente después de éste) los grupos permanecerán equivalentes en realidad. Fisher (1966) reconoció que la equivalencia alcanzada, independientemente de la habilidad experimental ejercida para lograrla, siempre será incompleta y, en algunos casos, "burdamente defectuosa".

Este problema existe, básicamente, debido al carácter discreto extremo de este fundamento. ¿Qué es lo que se está comparando?: un momento arbitrario, entre muchos otros, del comportamiento de los sujetos, bajo condiciones similares. Esta escasez de observaciones repetidas y su selección arbitraria determinan que las mediciones así obtenidas no sean estables ni sensibles, características *sine qua non* para una evaluación adecuada del tratamiento experimental.

En el caso de que se logre una equivalencia razonable entre los diferentes grupos, la siguiente operación en el diseño tradicional es la aplicación de distintos tratamientos a diferentes grupos. El razonamiento primordial es que, suponiendo que la conducta de los organismos de los distintos grupos fuese equivalente antes de los tratamientos y que éstos fueran efectivos, entonces el comportamiento de los organismos pertenecientes a un grupo debería ser netamente diferente del comportamiento de los organismos pertenecientes a otros grupos. Esta diferencia se determina de acuerdo a criterios estadísticos.

Cuando, de acuerdo a dichos criterios, se considera a una diferencia como importante, se hace una inferencia estadística que impugna y rechaza la hipótesis nula. En términos generales, la decisión estadística de rechazar o no la hipótesis nula estará bien fundamentada, siempre y cuando no se cometan los errores de tipo I o II. Recuérdese que una decisión estadística es de carácter probabilístico y que siempre existe la posibilidad de rechazar una hipótesis nula "verdadera" o de no rechazar una hipótesis nula "falsa".

Pero aun en el caso de que el procedimiento estadístico fuese impecable y la inferencia estadística fuera correcta, hay una segunda inferencia de tipo lógico atada a la inferencia estadística en el diseño tradicional. Su forma es, más o menos, la siguiente: si el control fue adecuado, si los grupos fueron inicialmente equivalentes, si éstos no cambiaron durante el experimento debido a fuentes de invalidez interna, si la medición tomada fue confiable y si la inferencia estadística

fue correcta, entonces la diferencia encontrada entre los grupos experimentales deberá indicar que hubo un efecto del tratamiento sobre la VD.

Cabe hacer notar que la inferencia lógica (la atribución de los cambios observados a los tratamientos experimentales) está múltiplemente condicionada; tienen que cumplirse varias clases de requisitos (condicionados a su vez a otros requisitos) en una forma *simultánea*, a fin de que la inferencia lógica se cumpla. En este caso, basta con que una sola de estas condiciones no se cumpla para derrumbar toda la estructura de la inferencia lógica experimental. Como las condiciones por cumplirse son muchas, es probable que un número sustancial de los experimentos señalados en la literatura hayan producido inferencias insostenibles ante el embate de esta crítica.

Aun en el caso de que todas estas dificultades hayan sido superadas y la inferencia lógica sea sólida, todavía hay tres objeciones metodológicas de importancia que hacer:

1. Los diseños tradicionales no contienen la posibilidad de *replicar directamente* los resultados obtenidos. Aun aquellos diseños que incluyen mediciones repetidas sobre los mismos sujetos son incapaces de proporcionar la estructura necesaria para hacer este tipo de replicación. Su formato solamente permite una apreciación vaga de los efectos de un tratamiento a través del tiempo, ya que se carece de un verdadero nivel de comparación como el de la línea base.

2. Los hallazgos logrados con grupos de sujetos no se pueden "extrapolar" directamente a la conducta de organismos individuales. Esto pudiera producir problemas realmente serios cuando se trata de hacer una explicación científica, así como en relación a aspectos de predicción y control.

3. Con el creciente uso de programas de computadora para el análisis de los datos, el investigador debe esperar a que sus datos sean codificados, perforados en tarjetas o grabados en cinta y finalmente analizados en la computadora, con el fin de saber si su procedimiento experimental fue efectivo o no. En contraste; un diseño con un solo sujeto (o un pequeño grupo de sujetos) medido múltiples veces de acuerdo a la tecnología de la línea base producirá datos que pueden analizarse inmediatamente, y el procedimiento experimental puede alterarse de acuerdo a la información obtenida de una manera simple (de tipo gráfico generalmente). Esto implica que un diseño tradicional complejo que produce múltiples datos puede dar lugar a una investigación relativamente lenta e insensible. Debe entenderse que se hace esta objeción no al uso de la computadora en sí (ésta sigue siendo el elemento más rápido para el procesamiento de grandes series de datos),

sino a la forma en que se ha establecido su uso. Por otra parte, cuando la computadora está en línea, analizando los datos conforme van ocurriendo, se le está dando el mejor uso posible, desde el punto de vista del experimentador.

10.2. EVALUACIÓN DE LOS DISEÑOS CONDUCTUALES

Para evaluar los diseños conductuales se utilizarán, básicamente, los mismos criterios que se usaron al estudiar los diseños tradicionales: *a*) las características de esta clase de diseños, y *b*) la validez de la aproximación metodológica que fundamenta a estos diseños.

10.2.1. Características de los diseños conductuales

Las características de este tipo de diseños también son las mismas que se estudiaron en la sección 10.1, a saber: *capacidad informativa, costo, poder, restricciones metodológicas, restricciones prácticas, propósito y efecto de la estadística.*

10.2.1.1. Capacidad informativa

Como sucede con los diseños tradicionales, los diseños más informativos son los más complejos; es decir, producirán más información aquellos diseños que tienen un mayor número de condiciones experimentales, una mayor cantidad de VI y un mayor número de VD.

Compárese, por ejemplo, un diseño A-B con un diseño A-B-A. Supóngase que, usando el primer diseño, se observará una diferencia notable en los niveles de la VD entre las fases A y B. Si ahora se añadiera la otra fase A del segundo diseño, éste incluiría la posibilidad de informar al investigador si el proceso bajo estudio es reversible o no. Esta información extra no puede conseguirse mediante un diseño A-B, y por esta razón se considera al diseño A-B-A como más informativo que el A-B.

Ahora, compárese un diseño A-B-A con uno A-B-A-B; la última fase 3 no solamente añade información relativa a la posibilidad de reincluir los efectos del tratamiento observados durante la primera fase B, sino que permite averiguar si es posible replicar el primer par de fases A-B; asimismo, permite hacer una replicación directa y produce un diseño más informativo que un diseño A-B-A.

Al aumentar el número de VI que se manipulan en un diseño del AEC, se incrementa su informatividad. Por tanto, un diseño multivariable será más informativo que uno univariable. Cuando se incluye más de una VI, se pueden intentar sus diferentes combinaciones, con el fin de indagar cuáles son los efectos de una VI por sí sola y cómo se ven aumentados o disminuidos por la inclusión de otra VI simultáneamente. Este tipo de información no es posible obtenerlo con diseños más simples. Finalmente, al aumentar el número de VD en un diseño, aumenta también su capacidad informativa. Un diseño A-B con línea base múltiple será tantas veces más informativo que un diseño A-B simple como líneas base incluya.

10.2.1.2. Costo

De acuerdo a lo estudiado en la sección 10.1, el costo de un diseño es igual al número de observaciones por diseño, y este número se obtiene multiplicando: *a*) el número de observaciones por unidad experimental; *b*) la cantidad de sujetos por unidad experimental; *c*) el número de condiciones experimentales, y *d*) la cantidad de VD. A continuación se comparan los costos de cuatro de los diseños conductuales más comunes.

Tabla 7

Costo (número de observaciones por diseño) de cuatro diferentes diseños conductuales.

Índice	DISEÑO			
	A-B	A-B-A	A-B-A-B	A-B × 4L.B.
Número de observaciones por unidad experimental	20	20	20	20
Número de sujetos por unidad experimental	1	1	1	1
Número de condiciones experimentales	2	3	4	2
Número de VD	1	1	1	4
Costo	40	60	80	

NOTA: A-B × 4L.B. = diseño A-B con línea base múltiple (cuatro líneas base).

Es importante hacer notar que el número de sujetos por unidad experimental será igual a 1 en la mayoría de los estudios dentro del AEC. También hay que señalar que el número de observaciones por unidad experimental (dentro de cada condición experimental) fue fijado arbitrariamente en 20, con el fin de simplificar la exposición. Nótese que los diseños más complejos no solamente son los más informativos, sino también los más costosos la mayoría de las veces.

10.2.1.3. Poder

Como ya se indicó, el poder de un diseño está en función de la relación entre su costo y su informatividad. Un diseño será más poderoso que otro cuando por el mismo costo genere más información. Esta relación se halla estrechamente asociada con el número de VI que se manipulan o con el número de VD que se pueden comparar.

Comoquiera que sea, dicha asociación no es una regla que siempre se cumpla, sino que depende de la comparación particular que se esté haciendo. Compárese, por ejemplo, un diseño A-B-A-B-A-B con uno A-B-A-B-A-C. Ambos diseños pueden contestar las mismas preguntas respecto al efecto que ejerce una VI sobre una VD, respecto a la naturaleza reversible o irreversible de dicho efecto y a la replicabilidad del cambio observado durante la secuencia A-B. Hasta este punto, ambos son equivalentes en su capacidad informativa. Ahora bien, como el diseño A-B-A-B-A-C incluye una nueva VI (fase C), puede contestar una pregunta respecto al efecto de ésta. Aunque el diseño A-B-A-B-A-B podría confirmar la primera replicación, esta información sería más bien redundante, mientras que su contraparte produciría una información completamente nueva. Además, en el segundo diseño se puede hacer una comparación entre los efectos de la primera VI y los efectos de la segunda. Ambos diseños tienen el mismo costo, pero el segundo es más informativo; por tanto, en este caso, el segundo diseño es más poderoso que el primero.

Ahora comparemos un diseño A-B-A-B con uno A-B-A-C. En este caso, no se puede decir que uno de ellos sea más informativo que el otro. Si bien el segundo de ellos contesta preguntas respecto a los efectos de dos VI sobre la misma VD y respecto a la reversibilidad, no da ninguna indicación respecto a la replicabilidad de la secuencia A-B. Por otra parte, el diseño A-B-A-B sí informa acerca de la replicabilidad y de la reversibilidad, pero solamente por lo que respecta a los efectos de una VI. En consecuencia, ambos diseños se consideran equivalentes en cuanto a su capacidad informativa. Si se supone, ade-

más, que ambos tienen el mismo costo, entonces se les podría atribuir el mismo poder. Este caso demostraría que la capacidad de un diseño de manipular más VI que otro, no lo hace necesariamente más informativo o más poderoso aunque éste pudiera ser el caso en múltiples investigaciones.

También el número de VD que incluye un diseño guarda una estrecha relación con su eficacia; por ejemplo, un diseño de línea base múltiple será generalmente más poderoso que un diseño equivalente con una sola línea base. Compárese, por ejemplo, un diseño A-B-A-B con uno A-B \times 2LB (con dos líneas base) y supóngase que ambos tienen un costo de 80 observaciones por diseño. De estos dos diseños con el mismo costo, probablemente el segundo sea el más informativo, ya que proporciona datos acerca del efecto que ejerce una VI sobre *dos* VD y permite una replicación con una diferente VD, sentando mejores bases para establecer la generalidad de la relación funcional encontrada (si hubiese alguna). Por su parte, el diseño A-B-A-B también permite una replicación, una reversión, e informa acerca de los efectos que ejerce una VI sobre una sola VD. La decisión última acerca de la informatividad (y, por tanto, del poder) de estos dos diseños depende de cómo se compara la información de la replicación en una segunda VD con la replicación directa en la misma VD, *más* la información acerca de la reversibilidad del fenómeno bajo estudio.

10.2.1.4. Restricciones metodológicas

Dos de las restricciones metodológicas de los diseños conductuales son similares a las de los diseños tradicionales: *a*) costo mínimo, y *b*) confiabilidad. Las otras son diferentes: *c*) inconsistencia de los criterios evaluativos, y *d*) uso repetido del mismo sujeto.

a) *Costo mínimo.* Esta característica es de discutible valor cuando se trata de diseños conductuales, ya que uno de los componentes del costo mínimo es el número mínimo de observaciones por unidad experimental y este número no se determina de acuerdo a reglas simples y teóricamente es imposible determinar de antemano (sobre todo, en la investigación básica). Por tanto, el concepto de costo mínimo resulta difícil de tratar.

Aquí sería útil enfatizar la diferencia entre el AEC en general y el ACA en particular. Aunque el primero se utiliza como un término genérico que abarca tanto los aspectos básicos como los aplicados al estudio de la conducta, en este contexto se le considerará como la rama dedicada a la investigación básica, y el segundo como la rama que

estudia la investigación aplicada. Esta diferencia en objetivos también conlleva diferencias en el rigor de la aproximación metodológica común y en la extensión de las restricciones metodológicas. El primer aspecto fue analizado brevemente en la sección 2.6. El segundo es fundamental para esta discusión. De tal suerte, si se ha de obtener una línea base estable, básicamente hay dos maneras de lograrlo: *a)* mediante el control directo de las variables, que a su vez controlan la conducta del organismo estudiado, como se hace en un laboratorio con animales, o *b)* esperando, después de múltiples observaciones, a que la línea base disminuya su variabilidad o, por lo menos, a que ésta se regularice. En ambas formas resulta difícil predecir con exactitud cuándo se estabilizará una línea base, especialmente en la última.

Por tanto, establecer el mínimo de observaciones por unidad experimental de antemano, está en contradicción con la aproximación metodológica sustentante del diseño; y sólo debido a las restricciones prácticas, en la investigación aplicada se fija un número mínimo en algunas ocasiones. Al respecto (Hall, 1971), indicó que dicho mínimo es igual a 5 y éste es el límite que se adopta en este libro, enfatizándose que *esto se indica bajo serias limitaciones de tipo práctico y que, en términos generales, el costo mínimo es indeterminado al iniciarse un estudio que implica el uso de un diseño conductual.*

b) Confiabilidad. Esta es una de las restricciones metodológicas características del AEC. Los datos de la conducta no sólo deben ser estables, sensibles y replicables, sino también, y en primer término, deben ser confiables. Cuando se tiene equipo automático de registro y éste ha sido debidamente calibrado y probado, los datos serán confiables; sin embargo, cuando no se tiene dicho equipo, se debe depender de los informes de observadores humanos. El grado de acuerdo entre éstos produce la medida de confiabilidad, y prácticamente todos los estudios realizados con la metodología operante en los cuales participan observadores humanos, informan de la confiabilidad de sus datos.

Aunque el concepto estadístico tradicional de confiabilidad es cuestionable y hay indicios de que el grado de acuerdo entre observadores independientes (concepto del AEC) no es la mejor medida de confiabilidad, el fundamento de estas mediciones o estimados se mantiene como una restricción metodológica irrefutable e indispensable. Si se han encontrado deficiencias a las formas en que se toma esta medición, lo único que se está indicando es que se desarrollen técnicas de medición más completas que realmente implementen el concepto de confiabilidad.

c) Inconsistencia de los criterios evaluativos. Los criterios utilizados para evaluar los cambios producidos con un diseño conductual

carecen de la uniformidad y el grado de formalización que tienen los criterios estadísticos. Cuando se emplea un diseño tradicional, generalmente no hay problema para evaluar la importancia de las diferencias observadas, pues ya hay una serie de reglas estadísticas que, si se siguen cuidadosamente, llevan a una decisión. Estas reglas son independientes de la experiencia del experimentador y dejan muy poco a su subjetividad, quizá solamente la elección de sus niveles de significancia. Por otra parte, cuando se usa un diseño conductual, quedan tres alternativas para resolver el problema de la evaluación de los datos: *a*) depender de la experiencia del investigador para hacer una apreciación visual subjetiva (analizando las gráficas de línea base y haciendo un juicio personal de su significancia); *b*) desarrollar nuevos medios cuantitativos para este propósito, como hizo Revushky (1967), o *c*) usar la inaceptable práctica de analizar los datos estadísticamente, como lo están haciendo numerosos investigadores en el área experimental.

La primera alternativa es sumamente común. No hay problema alguno al usarla, siempre y cuando el cambio producido entre una fase del diseño y otra sea realmente considerable. Si además se señalan los datos de la tasa de respuesta, se puede comparar el cambio visual con el cambio cuantitativo. El problema surge cuando los cambios no son considerables. Resulta muy difícil decidir si la diferencia es suficientemente "real". En estos casos limítrofes, sólo el investigador con experiencia podrá decidir con seguridad si los cambios logrados valen la pena; y si son demasiado dudosos, probablemente optará por alterar su procedimiento experimental, de tal manera que los cambios que produzca sean más obvios. Este asunto es controversial, ya que mientras el experimentador básico (con un criterio riguroso de lo que es el control experimental) argüiría que solamente se deben producir cambios tan notorios que hasta el investigador más inexperto los pueda advertir fácilmente, el experimentador con un entrenamiento cuantitativo argumentaría que los cambios deben evaluarse de acuerdo a criterios objetivos, independientemente del experimentador. Aún en la actualidad parece que no hay un acuerdo acerca de cuál de las dos posiciones se debe adoptar, a fin de obtener criterios evaluativos consistentes. Si bien la primera posición es aceptable desde el punto de vista del control experimental (más bien en la investigación básica), se sabe que hay determinadas situaciones en que el control no puede alcanzar la categoría del control del laboratorio, y entonces hay necesidad de recurrir a criterios cuantitativos que permitan discriminar entre una diferencia espuria y una verdadera.

La segunda alternativa parece promisoría: así como gran parte de la estadística y del diseño tradicional se desarrolló como una respuesta a los problemas planteados por la agricultura, también se puede desarrollar una “nueva estadística” (o un tipo diferente de análisis cuantitativo) para los datos conductuales. Ya hay numerosos esfuerzos en ese sentido, y cuando éstos fructifiquen la controversia anteriormente planteada se resolverá.

La tercera alternativa tampoco se debe perder de vista. Cuando se desconocen las suposiciones estadísticas fundamentales al uso de diversos análisis como correspondientes a los datos de la conducta que se están analizando, también se puede emplear la estadística. Como ya hay indicaciones de que algunas de las suposiciones básicas de la estadística pueden violarse, con cierta seguridad de que no se están produciendo inferencias deformes, el uso de una estadística sin suposiciones o restricciones podría demostrar ser de gran utilidad para el análisis de datos conductuales.

d) Uso repetido del mismo sujeto. Independientemente de las restricciones prácticas que lleva consigo el uso repetido del mismo sujeto, también hay varias restricciones de tipo metodológico que deben considerarse. En primer lugar, siempre existe la posibilidad de que se produzcan efectos residuales que se “acarrean” de una condición experimental a la siguiente. Estos efectos no permiten una evaluación adecuada del efecto de una VI o una verdadera recuperación de la línea base original.

En segundo lugar, hay la posibilidad de que se produzcan efectos “reactivos”, pero no en el sentido clásico de un problema de medición, sino en el de producir una “sobrerreversión”; es decir, cuando al incluirse una reversión a la línea base excede los límites de la línea base original (la tasa de respuesta es más alta). En tercer lugar queda la posibilidad (improbable) de producir un estado irreversible.

Todas estas alternativas imponen restricciones metodológicas al uso extensivo de un solo sujeto, uso típico en el diseño conductual, y deben considerarse como factores limitantes de dicha aproximación.

10.2.1.5. Restricciones prácticas

La selección de un diseño se ve generalmente afectada, y en ocasiones severamente limitada, por dificultades de tipo práctico como: *a)* el número de sujetos y su disponibilidad; *b)* el número de observadores independientes o experimentadores, y *c)* el presupuesto. A continuación se analizará cada una de ellas en relación al diseño conductual.

a) *Número de sujetos.* Este número no es una restricción práctica valedera tratándose del diseño conductual. Dado que no hay una exigencia metodológica de incluir un número mínimo de sujetos en un diseño, la mayoría de los experimentadores que emplean este tipo de diseño sólo incluyen unos cuantos sujetos, por lo general menos de diez y con cierta frecuencia solamente uno ($N=1$).

La otra restricción práctica relativa a los sujetos es su disponibilidad. Por lo común, cuando se trabaja en un laboratorio lo suficientemente grande y dotado de equipo y animales, éstos se hallan disponibles durante periodos tan extensos como un experimento lo requiera; sin embargo, cuando se trabaja con humanos, éste puede ser un problema serio.

Se sabe que una línea base puramente observacional debe contener (generalmente) un gran número de observaciones, las cuales deben presentar un rango pequeño de variabilidad para considerarse como verdaderamente estable. Ahora bien, no siempre es posible conseguir la colaboración de sujetos humanos durante periodos prolongados (véase la sección 10.1.1.5). El investigador aplicado se ve afectado particularmente por esta restricción y en muchas ocasiones tiene que sacrificar la calidad de su línea base, debido a la disponibilidad limitada de los sujetos humanos. En la investigación aplicada, entonces, en ocasiones se tiene que contraponer la restricción metodológica de obtener un número suficientemente grande de observaciones, a fin de alcanzar estabilidad en cada una de las etapas del diseño, contra la limitada disponibilidad temporal de los sujetos. Se podría afirmar que en múltiples investigaciones aplicadas, el número de sesiones experimentales es el resultado de un compromiso realista entre ambas exigencias incompatibles.

b) *Número de observadores o experimentadores.* El número de observadores independientes es una restricción práctica que aumenta con la cantidad de líneas base a registrar, el número de sujetos, etc. Esto resulta particularmente cierto cuando el diseño es complejo (por ejemplo, cuando contiene líneas base múltiples) o cuando la técnica de registro lo requiere (por ejemplo, cuando se registra por zonas o por individuos). Aunque no hubiera limitaciones presupuestarias, este requerimiento de contar con un número determinado de observadores entrenados puede limitar la complejidad de un diseño o de la técnica de registro. El número de experimentadores dependerá básicamente de cuántos sujetos se están utilizando y de la forma en que se aplican los tratamientos experimentales; así, a medida que aumentan ambos, se incrementa la necesidad de que haya más experimentadores, y esto pudiera afectar también la selección del diseño.

c) *Presupuesto.* El presupuesto disponible puede determinar la duración de un experimento, el número de sujetos, el número de observadores, etc. Conforme aumenta la cantidad de las mediciones tomadas, el número de sujetos, etc., aumentan los requerimientos de equipo, materiales, etc., y desde luego el costo monetario de un experimento. Como regla general, los diseños más complejos serán más costosos que los diseños más sencillos, y esto impone una restricción práctica importante al investigador que tiene la tarea de seleccionar o implementar un diseño.

10.2.1.6. Propósito

Hay dos propósitos fundamentales en el uso de un diseño conductual: a) a un nivel general, dictado por la estrategia de investigación característica del usuario de este tipo de diseño, está la contestación de preguntas de investigación, y b) a un nivel más específico, también dictado por la aproximación inductiva, se halla el establecimiento de relaciones funcionales entre diferentes variables que controlan la conducta (VI) y formas específicas de ésta (VD).

Ambos usos son prácticamente exclusivos con este tipo de diseño. Poco uso les es dado con el propósito de estimar parámetros y prácticamente ninguno para poner a prueba hipótesis. Cuando el propósito es el del establecimiento de relaciones funcionales, éstas se centran en los parámetros obtenidos con individuos particulares, y rara vez se plantean como producto de datos grupales.

10.2.1.7. Efecto de la estadística

En forma sumaria es posible asegurar que, hasta el presente, el efecto de la estadística sobre la metodología del AEC en general y sobre el diseño conductual en particular ha sido nulo. No hay diseños de esta clase que estén basados en o restringidos por suposiciones estadísticas. La cantidad de observaciones por sujeto o condición experimental, el número de sujetos por diseño, etc., son independientes de todo tipo de consideraciones estadísticas (aunque sería relativamente fácil relacionar algunos de ellos con la teoría de la medida). Finalmente, debe indicarse que el uso de la estadística para el análisis de datos producidos con esta clase de diseños está aumentando paulatinamente, aunque su importancia global es todavía insignificante.

10.2.2. Fundamento metodológico de los diseños conductuales

El fundamento metodológico en que se basa el diseño conductual se encuentra estrechamente relacionado con la estrategia inductiva de investigación y, en última instancia, con la posición atórica en psicología.

El investigador que sostiene esta posición no requiere de una teoría para hacer sus experimentos, ni tiene un interés particular en hacer predicciones (aunque pudiera hacerlas), ni le interesa poner a prueba hipótesis, ni otras conjeturas similares, y generalmente considera que aún en la actualidad no hay un cuerpo de datos relevantes lo suficientemente grande y consistente que sirva para empezar a formular una teoría. No obstante, está dispuesto a contribuir a la producción de dichos datos mediante el planteamiento de preguntas acerca del comportamiento y del diseño y conducción de uno o más experimentos para contestarlas, generando de esa manera más datos y más preguntas de investigación.

La contestación de dichas preguntas radica en el estudio analítico y exhaustivo de las relaciones funcionales entre la conducta (bajo control experimental) y sus determinantes. Para poder desarrollar dicho estudio, debe definirse objetivamente y sin ambigüedades cuál es la clase específica de respuesta a investigar, así como las condiciones antecedentes, concurrentes y consecuentes a su emisión. Cada uno de estos eventos debe ser analizado por separado y en su secuencia temporal. Una vez estudiadas algunas relaciones simples, se puede intentar investigar relaciones más complejas, incluyendo cada vez más eventos, lo que implica mejores técnicas de registro, control, medición y diseño.

El fundamento metodológico principal del diseño conductual es la línea base, es decir, la descripción objetiva y confiable de cada respuesta bajo estudio a través del tiempo. Dicha descripción tiene las siguientes restricciones: *a*) debe provenir del mismo sujeto o grupo de sujetos; *b*) debe representar adecuadamente la conducta "típica" del sujeto bajo las mismas condiciones, y *c*) debe ser capaz de advertir los cambios de la conducta cuando ocurran, así como indicar la magnitud de un cambio producido por una VI.

Uno de los fines de la estrategia inductiva es la generalidad de los datos, y ésta se obtiene a través de diferentes formas de replicabilidad. El diseño conductual contribuye a la consecución de dicho fin, proporcionando esquemas útiles para la replicación. Otro de los fines de esta estrategia es el eventual desarrollo de las técnicas de control de la conducta, y el diseño conductual también contribuye a su obten-

ción al permitir la evaluación de aquellos tratamientos que producen los cambios más considerables en los niveles de la(s) línea(s) base(s).

Por otra parte, el fin más importante para el investigador aplicado será la utilización de las técnicas de control conductual para la solución de problemas que implican la supresión de ciertas conductas, el mantenimiento de otras y la inclusión de otras más. El investigador aplicado también usa el diseño conductual para programar la secuencia de sus tratamientos y evaluar los cambios producidos.

Quizá sería prudente terminar el estudio de esta sección señalando las deficiencias de esta aproximación metodológica en general y del AEC en particular. En el siguiente capítulo se delinearán algunas proposiciones, a fin de superar dichas deficiencias.

La principal deficiencia del diseño conductual es la ausencia de criterios cuantitativos uniformes, independientes de la experiencia del experimentador. Esto es irónico, puesto que tal aproximación hace énfasis en lo objetivo y lo cuantitativo. Al rechazar a la estadística por una parte y a los modelos matemáticos por la otra, el científico conductual se ha colocado ante esta disyuntiva: o depende de criterios subjetivos (de su experiencia) o desarrolla criterios cuantitativos especialmente adecuados para este tipo de diseños.

Al presente, diferentes investigadores han optado por distintos métodos: una minoría se ha abocado a la tarea de desarrollar nuevas técnicas cuantitativas para el análisis de datos conductuales (Revushky, 1967). Otra minoría ha usado la rechazada estadística (Gentile y colaboradores, 1972). Otra tercera minoría se ha abocado a la tarea de hacer formulaciones matemáticas, a fin de presentar algunos hallazgos conductuales o evaluar el cambio (Hernstein, 1961, 1970; Rachlin, 1971; Killeen, 1972).

Otra deficiencia del diseño conductual es la escasez de niveles de comparación independientes (grupos control) para la evaluación de los datos; este punto se estudiará en el siguiente capítulo. Por último, otra deficiencia de esta aproximación metodológica (que afortunadamente ya ha sido reconocida y se está haciendo lo posible para corregirla) es la del uso casi exclusivo de la tasa de respuesta o algún índice que implique la frecuencia con que ocurre una respuesta. Aun cuando se admite que la tasa de respuesta es el dato básico de la conducta, esto no significa que sea el único. Hay datos como el tiempo interrespuestas, la latencia, la magnitud de la respuesta, la proporción de respuestas, etc., que se podrían utilizar más frecuentemente, como las VD del diseño conductual.

10.3. COMPARACIÓN ENTRE LOS DISEÑOS EXPERIMENTALES TRADICIONALES Y LOS DISEÑOS CONDUCTUALES

En vista de que ya en las secciones 10.1 y 10.2 se evaluaron los diseños de las clases tradicional y conductual, en esta sección solamente se incluyen los detalles sobresalientes de ambas, a fin de facilitar su comparación. Para hacer dicha comparación, se seguirá el mismo método empleado anteriormente: primero se compararán ambas clases de diseño de acuerdo a sus características más importantes, y después se compararán diversos aspectos de sus fundamentos metodológicos.

10.3.1. Características de las dos clases de diseño

Se compararán los diseños tradicionales con los diseños conductuales de acuerdo a las características estudiadas en las dos secciones previas.

10.3.1.1. Informatividad

No se puede afirmar, en forma general, que una clase de diseño sea más informativa que otra; esto depende de la complejidad de los diseños en cuestión, por una parte, y de la calidad del control experimental y de las mediciones tomadas, por otra.

10.3.1.2. Costo

El costo (en términos del número de observaciones) de dos diseños de diferentes clases debería estimarse en base a los datos específicos de ambos diseños. Aquí se seguirá el método de estudiar a cada uno de los componentes del costo de acuerdo al mínimo admisible de cada componente.

a) Número mínimo de observaciones por unidad experimental. En los diseños tradicionales, este número es igual a uno y en los diseños conductuales es indeterminado (como en la investigación básica) o es igual a cinco (como en la investigación aplicada). Este último mínimo es válido para cada fase de un diseño experimental dado, y no para la totalidad de las fases.

b) Número mínimo de sujetos por unidad experimental. En los diseños tradicionales, los mínimos aceptables se hallan entre diez y veinte

y rara vez se aceptan cinco. Por otro lado, esto no presenta ningún problema para los diseños conductuales, en los cuales el mínimo es igual a uno, ya que cada sujeto es considerado como una unidad experimental.

c) Número mínimo de condiciones experimentales. Este número es igual para ambas clases de diseño: dos. Esto es equivalente al mínimo de observaciones a que se refirió Campbell (1967).

d) El mínimo de VD. Este mínimo es, obviamente, igual a uno para ambas clases de diseño.

10.3.1.3. Poder

En términos generales, se puede decir que los diseños conductuales son más poderosos que los diseños tradicionales, aunque habría que hacer comparaciones específicas para ver hasta qué grado esta afirmación es válida.

Esta generalización está basada en un análisis de la forma en que organizan sus observaciones las dos clases de diseños. Los diseños tradicionales, básicamente, hacen comparaciones entre grupos (promedios de datos grupales), incluyendo a los diseños de grupos con mediciones repetidas. Estos diseños se pueden considerar como "cortes transversales" en un punto arbitrario de un "continuo" conductual. Los diseños conductuales fundamentalmente hacen comparaciones, utilizando al sujeto como su propio control (incluyendo aquellos casos en que además se comparan varios grupos). Estos diseños se podrían considerar como "cortes longitudinales" a lo largo de un determinado "continuo" conductual. Cuando se comparan dos ejemplos específicos de estas diferentes organizaciones (diseños) con el mismo costo, se puede observar que, a menudo, el diseño conductual resulta más informativo y, por tanto, más poderoso que su contraparte.

Así, en el extremo más simple, se puede comparar un diseño A-B con uno E-C, ambos con el mismo costo (por ejemplo, cuarenta). La diferencia principal es que mientras que el primer diseño incluye veinte observaciones en el mismo sujeto por cada fase del diseño, en el segundo se incluye una observación por cada sujeto en los dos diferentes grupos o condiciones experimentales.

Ambos diseños indicarían si hubiera un efecto de la VI sobre la VD y, hasta este punto, serían igualmente informativos. Ahora bien, el diseño A-B informaría, además, de qué manera se produce este efecto en observaciones sucesivas, indicando si éste es gradual o repentino, inmediato o retardado. Por otra parte, sin embargo, el diseño E-C no puede producir este tipo de información, y los datos relativos a un sujeto cualquiera en particular serían irrelevantes.

Incluso, se podría argüir en favor del diseño tradicional, señalando que éste provee información representativa de una determinada población y que, por tanto, da indicaciones que tienden hacia la generalidad. Este argumento es endeble en dos aspectos: *a*) a menos que el proceso de muestreo que produjo los grupos en cuestión hubiese sido técnicamente nítido, es difícil clamar tal representatividad, y *b*) aun cuando esto se lograra, tener información acerca de los individuos integrantes de un grupo en el momento x en que se midió la VD, no garantiza que haya la posibilidad de que algunos (o todos) de ellos se comporten en una forma totalmente diferente en el momento y o z , lo cual hace cuestionable a los datos obtenidos en el momento x y a los razonamientos basados en ellos. Vemos, así, que aun cuando la información producida por los diseños tradicionales es, en apariencia, equivalente a la producida con los diseños conductuales, está condicionada al cumplimiento de ciertas suposiciones, haciendo a dichos diseños comparativamente menos poderosos.

10.3.1.4. Restricciones metodológicas

Las restricciones metodológicas en las que se compararán las dos clases de diseños son: el *costo mínimo*, la *confiabilidad*, el *uso de grupos*, la *utilización de individuos*, el *empleo de la estadística*, el uso de las *escalas de medición* y los *criterios evaluativos*. A continuación se estudia cada uno de ellos.

a) Costo mínimo. Por lo general, los diseños tradicionales son aproximadamente tan costosos como los diseños conductuales (quizá con la única excepción de estudios masivos en los cuales se incluyen múltiples sujetos por cada grupo). Sin embargo, como el costo mínimo de los diseños conductuales utilizados en la investigación básica es indeterminado pero tiende a requerir un gran número de observaciones por sujeto, frecuentemente éstos pueden ser más costosos que los diseños tradicionales.

b) Confiabilidad. La importancia que se le da a esta medición es muy diferente en ambas clases de aproximación metodológica. Mientras en los diseños conductuales el reporte de la confiabilidad es un requisito indispensable, en los diseños tradicionales no se hace tanto énfasis en este aspecto y la confiabilidad de las mediciones se señala relativamente con poca frecuencia. También la forma en que se considera a la confiabilidad es diferente. Mientras en el diseño tradicional es una correlación (r) entre dos o más mediciones sucesivas o paralelas, en el diseño conductual es la proporción de acuerdo (grado de acuerdo) entre dos o más observadores independientes.

c) *Uso de grupos.* Los grupos de sujetos constituyen las unidades experimentales en los diseños tradicionales, y las comparaciones que se hacen entre ellos son la base misma de esta clase de diseños. Por otra parte, el uso de grupos es de importancia secundaria para los diseños conductuales. Aunque en este tipo de diseños se enfatizan los cambios de conducta en el mismo sujeto, en ocasiones se agrupan los datos de sujetos individuales e, incluso, se hacen experimentos con uno o más grupos, en los cuales se toman mediciones repetidas y se comparan los grupos. Krantz (1971) hizo una comparación de las prácticas editoriales de dos revistas técnicas, representando a las dos aproximaciones metodológicas de nuestro estudio: el JEAB (*Journal of the Experimental Analysis of Behavior*) que representaba al AEC, y el JCPP (*Journal of Comparative and Physiological Psychology*), que representaba a la metodología tradicional. Esta comparación la hizo con los números de los años 1967-1969 y encontró que en el JEAB se publicaron entre el 87.5% y el 91.2% de estudios utilizando al mismo sujeto como su propio control, contra sólo entre el 1.8% y el 3.8% comparando grupos de sujetos. Por su parte, el JCPP publicó el 88.5% de estudios, haciendo una comparación entre grupos, contra sólo entre el 6.0% y el 6.6% con el mismo sujeto como su propio control.

d) *Utilización de individuos.* Esta es la práctica más indicada por los diseños conductuales y la que se sigue en la mayoría de los casos, bajo la aproximación metodológica operante. Por lo contrario, el análisis de los datos de un solo individuo en el diseño tradicional es, en la mayoría de las veces, inútil.

e) *Empleo de la estadística.* Este punto es fundamental para diferenciar a los diseños tradicionales de los diseños conductuales. Como ya se ha señalado antes, no es raro que se identifique a los primeros como "diseños estadísticos". En contraste, el empleo de la estadística para el análisis de los datos del diseño conductual es sumamente limitado.

f) *Escalas de medición.* Una regla con amplia aceptación entre los usuarios del diseño tradicional es que los datos deben corresponder por lo menos a una escala de intervalos (la cual supone que la distancia entre un punto y el siguiente en la escala forma un intervalo equivalente al que se pueda formar con cualquier otro par de puntos sucesivos) o a una escala de proporción (la cual corresponde a escalas físicas, como la de la temperatura, longitud, etc.), si fuese posible. Los diseños conductuales, por otra parte, no hacen restricciones metodológicas basadas en las escalas de medición.

g) *Criterios evaluativos.* Aquí la diferencia es a favor de los diseños tradicionales. Ya que los criterios para la evaluación de los da-

tos siempre son de tipo estadístico y por estar basados en las mismas suposiciones y seguir primordialmente la misma secuencia o estructura lógica, son consistentes entre sí. Por otra parte, los criterios del AEC carecen de esa uniformidad, pues no hay un consenso general respecto a la forma más objetiva y cuantitativa de evaluar los datos que proporcione reglas independientes del investigador para dicho propósito.

10.3.1.5. Restricciones prácticas

Las restricciones prácticas en que se compararán las dos clases de diseño son: *sujetos, observadores o experimentadores, y presupuesto.*

a) *Sujetos.* En este punto hay un balance entre ambos tipos de diseños. Los diseños tradicionales generalmente exigen números moderados (20-50) o grandes (100-500) de sujetos por grupo, y los diseños conductuales se pueden implementar con un solo sujeto. En contraposición, los diseños conductuales requieren un número moderado (10-50) o grande (100- n) de mediciones repetidas por unidad experimental dentro de una fase del diseño, y los diseños tradicionales pueden requerir una o dos de dichas mediciones. En este sentido, se entiende que hay un balance entre ambas clases de diseño. Obviamente, para algunos experimentadores resultará más difícil vencer la dificultad de conseguir un gran número de sujetos, mientras que para otros lo será el tener acceso a ellos por periodos prolongados.

b) *Número de observadores o experimentadores.* En ambos casos, este número variará de acuerdo a la complejidad del diseño y a la cantidad de VD a registrarse; sin embargo, en aquellos casos en que no se cuenta con un registro automático de medición, los diseños conductuales requieren cuando menos dos observadores independientes para informar de la confiabilidad de los datos, lo cual haría más difícil la satisfacción de esta restricción práctica.

c) *Presupuesto.* En este punto sería imprudente comparar ambas clases de diseños en general y hacer una afirmación global, indicando que una de ellas es más costosa (en términos monetarios) que la otra. El presupuesto de un experimento dependerá de una serie de factores relacionados con el diseño experimental, pero tal relación no es simple. El presupuesto estará en relación al "costo" (en términos de observaciones) del diseño, al número de observadores o experimentadores, al tipo de sujeto, al equipo de estimulación o medición, al pago o mantenimiento de sujetos, etc.

10.3.1.6. Propósito

En una forma simplificada, se puede decir que el propósito principal de los diseños tradicionales es el de poner a prueba diferentes hipótesis. Un propósito secundario es el de indicar relaciones funcionales de datos grupales. Por su parte, los propósitos fundamentales del diseño conductual son el establecimiento de relaciones funcionales (que empiezan a nivel individual y se extienden hasta representar a una o más especies) y la contestación de preguntas de investigación. Ambas clases de diseño tienen, escasamente, el propósito de estimar parámetros.

10.3.1.7. Efecto de la estadística

En los términos más simples, se puede decir que la estadística ha contribuido a moldear (si no es que a determinar) al diseño tradicional. Esto lo ha hecho imponiéndole restricciones dirigidas a satisfacer ciertas suposiciones de la teoría estadística, indicando un proceso inductivo basado en la inferencia estadística, e influyendo en la estrategia de investigación mediante el énfasis en la prueba de hipótesis. Se puede decir, que el efecto de la estadística sobre el diseño conductual es mínimo o nulo en lo que se refiere a la estructura de esta clase de diseño y a sus procesos inductivos. Quizá se debiera añadir que ha ejercido algún efecto indirecto sobre los usuarios y creadores del diseño conductual, para producir técnicas cuantitativas alternativas.

10.3.2. Otras características

Además de las características de los diseños estudiadas en las secciones de evaluación (10.1 y 10.2), aquí se incluirán otras características que son importantes para efectos de desarrollar una comparación más exhaustiva. Así, tenemos: a) el *rango secuencial de las observaciones*; b) la *replicabilidad*; c) la *capacidad multivariable*; d) los *niveles de comparación*; e) la *reorganización de las unidades experimentales*, y f) la *protección contra las fuentes de invalidez*. A continuación se analizan brevemente.

10.3.2.1. Rango secuencial de las observaciones

Los diseños conductuales generalmente tienen un rango de observaciones mucho mayor que el de los diseños tradicionales. Esto es

válido para cada fase de la secuencia: antes, durante y después de la aplicación de un determinado tratamiento experimental. Los diseños tradicionales sólo en algunas ocasiones incluyen observaciones previas al tratamiento (por ejemplo, mediante pretests) o durante el tratamiento (algunos casos de mediciones repetidas). El número de observaciones posteriores al tratamiento también es menor al de los diseños conductuales (aun cuando hay investigación preexperimental de tipo longitudinal en la cual el seguimiento de los sujetos ha durado hasta diez o más años).

10.3.2.2. Replicabilidad

Aquí haremos una diferenciación entre dos posibilidades de replicación: aquella determinada por la estructura misma de un diseño y que denominaremos como 'autocontenida' (o intrínseca) y aquella independiente de la estructura del diseño o "extrínseca". Esta última posibilidad no será considerada como importante, ya que es factible en ambas clases de diseño.

La primera posibilidad, sin embargo, establece una importante diferencia entre los diseños tradicionales y los conductuales: los últimos ofrecen la posibilidad de hacer una replicación directa (y en algunos casos una replicación sistemática) como parte de su estructura; se dice, entonces, que la replicabilidad está autocontenida. Dicha posibilidad se halla ausente en la mayoría de los diseños tradicionales (con excepción de los diseños replicados que cita Winer, 1962), incluso en aquellos con mediciones repetidas.

Un ejemplo de replicabilidad autocontenida está dado por un diseño A-B-A-B; otro, por un diseño de línea base múltiple. En ambos existe la posibilidad de demostrar repetidamente los efectos que ejerce la VI sobre la(s) VD.

10.3.2.3. Capacidad multivariable

Esta característica se refiere a la posibilidad de contener múltiples variables en un diseño. Es importante hacer la distinción entre dos tipos de capacidad multivariable: a) capacidad multivariable manipulativa, y b) capacidad multivariable de medición. La primera obedece al sentido que se le ha dado al término *multivariable* en este libro, o sea, el de incluir para su manejo múltiples VI. La segunda obedece al significado más tradicional en boga del término *multivariable*, que es el

sentido estadístico o de medición, o sea, aquel que comprende el registro o procesamiento de datos de múltiples *VD*.

En ambos tipos de diseño la capacidad multivariable manipulativa va desde los diseños simples, que son univariantes, hasta los diseños complejos, que son multivariantes. Sobre el papel se pueden construir diseños multivariantes con una gran capacidad manipulativa tanto con diseños de un tipo como del otro. En este punto se puede decir que no hay superioridad de uno u otro tipo de diseños. Por lo que respecta a la capacidad multivariable de medición hay una definitiva superioridad de los diseños tradicionales sobre los conductuales. Prácticamente todos los diseños tradicionales permiten la medición simultánea de varias *VD*, mientras que sólo algunos de los diseños conductuales permiten lo mismo.

10.3.2.4. Niveles de comparación

La diferencia entre ambos tipos de diseños en este respecto es que, mientras en la mayoría de los diseños tradicionales los niveles de comparación (grupos) son independientes entre sí, en la mayoría de los diseños conductuales no lo son. En el caso de los diseños tradicionales, generalmente se incluyen uno o más grupos control como el nivel de comparación básico; y aun en el caso de que no se incluya un grupo control, siempre hay otros grupos experimentales que sirven como niveles de comparación independientes. La independencia radica en que, con algunas excepciones, ningún sujeto puede servir en dos o más condiciones; de manera que lo que se produce son observaciones estadísticamente independientes. Por otra parte, ya que en los diseños conductuales cada sujeto sirve como su propio control, esta independencia de las observaciones no se cumple. Cabe añadir que en algunos casos se incluye algún grupo control que tiene la función de un nivel independiente.

10.3.2.5. Reorganización de las unidades experimentales

Cuando se trabaja con un diseño conductual es posible reunir las diferentes unidades experimentales (individuos) y crear un grupo que puede ser considerado también como una unidad experimental en sí; es decir, independientemente del análisis que se haga de los datos de cada individuo en particular, también se les puede reorganizar como un grupo. En el caso opuesto, en que el interés de un experimentador

esté dirigido a un grupo de sujetos, siempre tiene la posibilidad de dividirlo en sus componentes individuales y considerarlos como unidades experimentales independientes. En los diseños tradicionales esto no es posible, puesto que se analizan los datos del grupo y no los de un individuo en particular.

10.3.2.6. Protección contra las fuentes de invalidez

En los capítulos 2 y 3 se estudiaron dos posibles fuentes de invalidez: la interna y la externa. A continuación veamos cómo difieren los diseños tradicionales de los diseños conductuales en la forma de protegerse contra estas fuentes de invalidez.

a) *Fuentes de invalidez interna.* Los diseños tradicionales utilizan un cierto número de medidas para controlar o minimizar los riesgos de dichas fuentes de invalidez. Una de ellas es la aleatorización con la cual se minimizan los efectos de la selección. La otra es el muestreo que, cuando se usa adecuadamente, puede disminuir los efectos de la mortalidad. Los efectos de la "historia", la maduración y la regresión estadística se evitan cuando no se toman mediciones repetidas, o se reducen cuando no se deja transcurrir mucho tiempo entre una y otra medición. Los problemas de la medición se intentan reducir mediante estudios de confiabilidad de los llamados "instrumentos de medición" (pruebas psicológicas). Otras formas de invalidez interna, como la instrumentación (a excepción de sus efectos reactivos) o la interacción, exceden las posibilidades de control a través del diseño o de su implementación.

Por otra parte, las fuentes de invalidez interna se pueden considerar, en su mayoría, como irrelevantes para los diseños conductuales, puesto que su control está implícito en la estructura de este tipo de diseño; por ejemplo, la línea base conductual resuelve automáticamente los problemas de "historia", creando una "historia reciente" bajo el control experimental; el de la maduración, ya que cualquier cambio importante puede ser advertido y sólo los que son producto de la manipulación experimental se repetirán bajo una de las formas de replicación; y el de la regresión estadística, ya que este problema pierde su fuerza cuando hay un gran número de observaciones sobre el mismo sujeto. El problema de medición se reduce por medio del registro automático o mediante el aumento de la confiabilidad interobservadores. La "mortalidad" es un problema de cierta importancia cuando $N=1$, en cuyo caso se tiene que repetir completamente el experimento con un nuevo sujeto. Cuando hay más de un sujeto, la pérdida de uno

de ellos no constituye un problema serio, ya que al reducirse N no se afecta realmente el significado de los datos. Los problemas de instrumentación o de interacción también exceden esta clase de diseños, pero su presencia podrá notarse más fácilmente debido al uso extensivo de la instrumentación a lo largo del diseño y al registro continuo de los datos.

El elemento más eficaz para resolver los problemas de invalidez interna es el control experimental. Un buen manejo de las variables relevantes, junto con un diseño flexible y eficaz, pueden disminuir sustancialmente este tipo de invalidez. Finalmente, un medio efectivo para obtener validez interna al alcance del investigador empleando una u otra clase de diseño es la replicación sistemática. Si un resultado se replica una y otra vez bajo circunstancias semejantes (en lo que respecta a las variables implicadas) pero diferentes (en lo que se refiere a los procedimientos), entonces se puede tener bastante seguridad de que el efecto de las fuentes de invalidez interna fue despreciable.

b) *Fuentes de invalidez externa.* Como se planteó en el capítulo 2, la validez externa de un experimento depende considerablemente de que los datos reflejen en una forma inalterada los cambios ocurridos en algún índice relacionado con el comportamiento del organismo estudiado; es decir, debe evitarse que el cambio observado sea (aunque parcialmente) producto del proceso de medición. Cuando esto se logra, lo cual no es fácil, se producen mediciones no reactivas y los datos son muy valiosos. Dado que los datos corresponden a cambios reales en un organismo o en su comportamiento, entonces resulta importante saber si son representativos de otros organismos de la misma especie. Si lo son, generalmente tendrán más valor que si sólo correspondieran a un caso raro de los estudiados en la sección 4.5. Finalmente, se tratará de extender el alcance de los resultados de un hallazgo experimental, de ser posible, a otras especies. A esto se le denomina *generalidad de los datos a través de la replicación interespecie* y es una característica que permite establecer a un fenómeno como de importancia "universal" y en la cual se basa la credibilidad de un hallazgo original. A continuación se compararán las dos clases de diseños en este aspecto.

Los diseños tradicionales emplean, básicamente, la aleatorización y las técnicas de muestreo (ambas con una orientación estadística) como medios de protección contra las fuentes de invalidez externa. Los investigadores que usan este tipo de diseño emplean alguna forma de replicación y ocasionalmente alguna forma de observación no obstrusiva

(que no es percibida por los observados) como una forma de evitar la invalidez externa.

Los investigadores que usan los diseños conductuales dependen fundamentalmente de la replicabilidad de sus resultados para aumentar la generabilidad de sus hallazgos. El problema de reactividad de las mediciones se resuelve por medio del uso de periodos de adaptación generalmente contenidos en los diseños y con la ayuda de equipo automático de registro que, por supuesto, es independiente del diseño. El problema de la representatividad, en términos de muestreo, es de importancia menor cuando se emplea un diseño conductual, siempre que los datos sean confiables y replicables.

La tabla 8 presenta en una forma concisa los puntos de comparación entre los diseños tradicionales y los conductuales que se estudiaron en este capítulo.

Índice de comparación	CARACTERÍSTICAS				
	Informatividad	Poder	Rango secuencial de las observaciones		
			Observaciones antes del tratamiento experimental	Observaciones durante el tratamiento experimental	Observaciones después del tratamiento experimental
Clase de diseño					
Diseños tradicionales	Variable. Aumenta con la complejidad del diseño	Variable. Depende de la relación costo-informatividad	Generalmente entre 0-1	Generalmente 0	Generalmente entre 1-20
Diseños conductuales	Variable. Aumenta con la complejidad del diseño	Variable. Depende de la relación costo-informatividad. En general, mayor que el del diseño tradicional. Produce relaciones más firmes	Número claramente mayor que en el diseño tradicional. 5-n	Número claramente mayor que en el diseño tradicional 5-n	Generalmente mayor que en el diseño tradicional. 5-n
Índice de comparación	RESTRICCIONES METODOLÓGICAS				
Clase de diseño	Protección contra las fuentes de invalidez externa	Costo mínimo: número mínimo de observaciones por unidad experimental	Costo mínimo: número de sujetos por unidad experimental (N)	Costo mínimo: número mínimo de condiciones experimentales	Costo mínimo: número mínimo de variables dependientes
Diseños tradicionales	Uso ocasional de medición no reactiva y replicación directa o sistemática. Muestreo	1	10-20 (cinco en casos excepcionales)	2	1
Diseños conductuales	Uso extenso de replicación directa y sistemática. Adaptación	Cinco en el ACA (por cada fase). Indeterminado en el AEC (básico)	1	2	1
Índice de comparación	RESTRICCIONES PRÁCTICAS				
Clase de diseño	Criterios evaluativos	Sujetos	Número de observadores o experimentadores	Presupuesto	Prueba de hipótesis
Diseños tradicionales	Uniformes y objetivos (estadísticos)	Número: véase costo mínimo. Disponibilidad: se pueden usar en casos de disponibilidad irrestringida	Variable. Dependiente del experimento	Variable. Dependiente de la complejidad del diseño y del procedimiento. Puede determinar la selección del diseño	Uno de los propósitos fundamentales
Diseños conductuales	Inconsistentes. Generalmente objetivos. Ocasionalmente subjetivos	Número: véase costo mínimo. Disponibilidad: no se pueden usar en casos de disponibilidad restringida	Variable. Dependiente del experimento. Cuando no se usa equipo automático de registro se requieren por lo menos dos observadores	Variable. Dependiente de la complejidad del diseño y del procedimiento. Puede determinar la selección del diseño	No se intenta

CARACTERÍSTICAS

Capacidad	Multivariable	Niveles de comparación	Reorganización de las unidades experimentales	Protección contra las fuentes de invalidez interna	
Manipulativa	De medición				
Contenida: Conte	Variable. Depende de si el diseño es univariable, bivariable o multivariable	Permiten la medición simultánea de varias VD	Generalmente independientes de las unidades experimentales (otros grupos de sujetos)	No es posible hacer particiones ni reagrupaciones	Aleatorización, muestreo, confiabilidad y control experimental
Contenida: Conte	Variable. Depende de si el diseño es univariable, bivariable o multivariable	Sólo unos cuantos permiten la medición simultánea de varias VD	Dependientes de las unidades experimentales (cada sujeto es su propio control)	Es posible reagrupar las unidades experimentales y también hacer particiones de grupos	Está implícita en la tecnología de la línea base. Control experimental

RESTRICCIONES METODOLÓGICAS

o mínimo: ero mínimo observadores diseño	Confiabilidad	Uso de grupos	Uso de organismos individuales	Uso de la estadística	Uso de las escalas de medición
20-40	No es un requisito indispensable. Se le reporta con poca frecuencia	Indicado por la mayoría de estos diseños	Escaso	Prácticamente imprescindible. Todos los análisis de datos son de tipo estadístico	Los datos deben corresponder cuando menos a una escala de intervalo
en el ACA terminado en EC (básico)	Es un requisito indispensable cuando no hay equipo automático de registro	Posible, pero no indispensable	Indicado por estos diseños	Se le considera de importancia secundaria o irrelevante. Uso limitado	No hay restricciones al respecto

PROPÓSITO

gunta de in- gación	Establecimiento de relaciones funcionales	Estimación de parámetros	Estrategia de investigación	Efecto de los modelos estadísticos	Uso en la investigación psicológica
limitado	Grupales. Generalización a conducta individual riesgosa	Uso muy limitado	Generalmente deductiva	Importante	Frecuente
de los pro- tos funda- tales	Se logra en forma inequívoca (parámetros limitados a los sujetos bajo estudio)	Uso muy limitado	Inductiva	Nulo	Frecuente. En aumento

11

compatibilidad entre los diseños conductuales y los diseños tradicionales

EN EL CAPÍTULO anterior se discutieron exhaustivamente aquellos aspectos importantes para la comparación, evaluación y selección de diseños experimentales en la ciencia de la conducta y disciplinas afines. Aun cuando se compararon los diseños conductuales con los tradicionales y se señalaron ejemplos mostrando superioridad los de una u otra clase en aspectos específicos, no se contempló la posibilidad de que fuesen compatibles entre sí. En este capítulo se analizarán tres aspectos en los cuales pudiera haber compatibilidad o incompatibilidad entre los diseños conductuales y los tradicionales: *a)* metodología; *b)* estructura, y *c)* evaluación o análisis de los datos. También se estudiará la posibilidad de desarrollar algunos nuevos diseños que podrían demostrar ser de utilidad en la investigación conductual.

11.1. METODOLOGÍA

En el capítulo anterior quedó implícita la incompatibilidad entre la aproximación metodológica conductual y la tradicional; ahora se le formula explícitamente. La metodología en la que se fundamentan los diseños experimentales tradicionales es incompatible con la que fun-

damenta a los diseños experimentales conductuales, por las siguientes razones (entre otras): *a*) la estrategia de investigación típica del diseño tradicional está dirigida hacia la prueba de hipótesis (recuérdese que una hipótesis es una afirmación) mientras que la estrategia de investigación del diseño conductual está dirigida hacia la contestación de preguntas de investigación; *b*) mientras las relaciones funcionales encontradas con el diseño tradicional son, en su mayor parte, de tipo grupal, las del diseño conductual son tanto de tipo individual como de tipo grupal, primordialmente del primero; *c*) los supuestos básicos y los procedimientos experimentales que de ellos se derivan también son diferentes: mientras en el diseño tradicional se procede a "igualar" a los grupos de sujetos, en el conductual se trata de obtener una línea base; *d*) mientras en la aproximación tradicional se busca obtener un gran número de observaciones aportadas por múltiples individuos en un momento arbitrario, en la segunda se busca una gran cantidad de observaciones aportadas por un solo organismo en momentos sucesivos (la clásica discusión "grupos vs organismos individuales"); *e*) la inferencia estadística es indispensable en el diseño tradicional y no lo es en absoluto en el conductual; *f*) el problema de la variabilidad se trata en términos del control experimental en el AEC y en términos de la medición en la aproximación tradicional (en el primer caso se le manipula y en el segundo "se le respeta"). También hay diferencias notables en el rango secuencial de observaciones, en la influencia de la estadística y en los criterios evaluativos.

Es evidente que estas diferencias revelan una incompatibilidad fundamental entre ambas aproximaciones: no es compatible el iniciar una investigación con una hipótesis (una afirmación) con iniciarla mediante una pregunta, ni se puede hacer una equivalencia entre múltiples mediciones de un solo sujeto y una sola medición de múltiples sujetos, etcétera.

A estas alturas seguramente el lector ya tiene suficientes elementos de juicio como para hacer una evaluación de ambos tipos de metodología. También el autor los tiene y su opinión personal es la siguiente: dado que: *a*) los diseños conductuales tienen un mayor poder (comúnmente) que los diseños tradicionales, *b*) su estructura propicia un mayor rango secuencial de observaciones, *c*) hacen factibles las principales formas conocidas de replicación, *d*) permiten la reorganización de las unidades experimentales, *e*) los diseños conductuales ofrecen mejores medidas de protección contra las fuentes de invalidez que su contraparte tradicional, *f*) hacen énfasis en la confiabilidad de los datos, *g*) la metodología conductual estudia exhaustivamente a los organismos comportantes bajo condiciones de control, y *h*) sus resultados

tienden a ser más precisos e inequívocos, considera que esta aproximación es superior a la tradicional. Esta opinión, desde luego, puede ser impugnada por el lector y lo que la pondrá a prueba será, en última instancia, la tendencia que tome la investigación conductual (psicológica en general) en lo futuro y los logros que se obtengan con cada una de las dos aproximaciones metodológicas.

11.2. ESTRUCTURA

Independientemente de la manifiesta incompatibilidad entre ambas aproximaciones metodológicas, se puede observar, sin mucha sorpresa, que las estructuras de los diseños tradicionales son compatibles con las de los conductuales. Es posible, por ejemplo, tener un diseño que contenga un grupo experimental y un grupo control (estructura tradicional) y al mismo tiempo se puede estar siguiendo la reversión A-B-A (estructura conductual). Ésta sería una de las muchas combinaciones que es posible formar con estructuras independientes. En la literatura contemporánea es posible encontrar un número de combinaciones de este tipo; por ejemplo, Burchard y Barrera (1972) estudiaron los efectos del tiempo-fuera de reforzamiento y el costo de la respuesta sobre la conducta antisocial de sujetos retardados. Informaron haber empleado un diseño de cuatro condiciones (grupos), pero además usaron una secuencia A-B; es decir, los sujetos fueron observados durante la fase A, antes de la aplicación de las diferentes combinaciones de las VI (fase B). Las cuatro combinaciones de las VI tiempo-fuera (de reforzamiento) y costo que formaron los cuatro grupos fueron, en tiempo y fichas, respectivamente: 0-5, 0-30, 5-0 y 30-0. En todas estas condiciones se requirieron doce días de línea base durante la fase A. Otra investigación en la que se utilizaron cuatro grupos y mediciones repetidas A-B fue realizado por Marston y McFall (1971). Otras combinaciones son: *a*) grupo control-grupo experimental y mediciones repetidas A-B, que es, aparentemente, la más común [Bensberg y colaboradores (1965), McKenzie y colaboradores (1968), Resnick (1968), Lovitt y colaboradores (1969), Schwitzgebel (1969), McNamara (1971) y Baker (1971)], y *b*) grupo control-grupo experimental y diseño reversible A-B-A (Ryan y Krumboltz, 1964).

Hay más combinaciones posibles de diseños tradicionales con diseños conductuales. La tabla 9 presenta sólo algunas de las más factibles, a juzgar por la frecuencia con que se usan los diseños independientes que forman las combinaciones.

Tabla 9

Algunas de las combinaciones más factibles de diseños tradicionales con diseños conductuales

		DISEÑOS CONDUCTUALES					
		AB	ABA	ABAB	AB- -AB	ABCA	ABC- -X-Y-A
Diseños tradicionales	E-C	+	+	+	+	++	++
	E-C ₁ -C ₂	+	+	+	+	++	++
	C-E ₁ -E ₂	+	+	+	+	++	++
	C-E ₁ - -E _n	+	+	+	+	++	++
	E-C ₁ - -C ₂	+	+	+	+	++	++
	2 × 2	++	++	++	++	+++	+++
	n × n	++	++	++	++	+++	+++
	2 × 2 × 2	++	++	++	++	+++	+++
	n × n × - - × n	++	—	—	—	—	—

NOTA: léase la primera celdilla como EC × AB, la segunda como EC × ABA, la última como (n × n × - - × n) × (ABC- -XY-A), etc.

Aunque es factible llevar a cabo experimentos que impliquen cualquiera de estas combinaciones, hay algunas de ellas que serían realmente prohibitivas en términos del número de observaciones; éstas se han señalado con el símbolo “menos” (—). Las opciones que permiten la combinación de diseños multivariantes de las dos clases y que, a pesar de que resultan costosas, no son prohibitivas, están señaladas con el símbolo +++. Aquellas celdillas que corresponden a una combinación de un diseño multivariable con un diseño más simple aparecen marcadas con el símbolo ++. El resto de las combinaciones incluye los niveles más simples (univariantes o bivariantes) de diseños de ambas clases y, en su mayoría, son del tipo de combinación ya señalado en la literatura. Dichas combinaciones aparecen marcadas con el símbolo +.

Algunos de los diseños resultantes de estas combinaciones pueden ser verdaderamente poderosos (muy informativos a un costo razonable). Considérese, por ejemplo, un diseño 2 × 2 × ABA con un solo sujeto por cada grupo. Dicho diseño tendría doce condiciones experimentales, en vez de las cuatro o tres que tiene cada uno de sus diseños componentes por separado. Este diseño implica el uso de tres diferentes VI y la participación de cuatro sujetos en total. Supóngase que se toman diez mediciones por cada fase; de esta manera se obtendrían 2 × 2 × 3 × 10 = 120 observaciones (30 por sujeto). Una representación tridimensional de este diseño se encuentra en la figura 35 y una representación bidimensional de los mismos datos se halla en la figura 36.

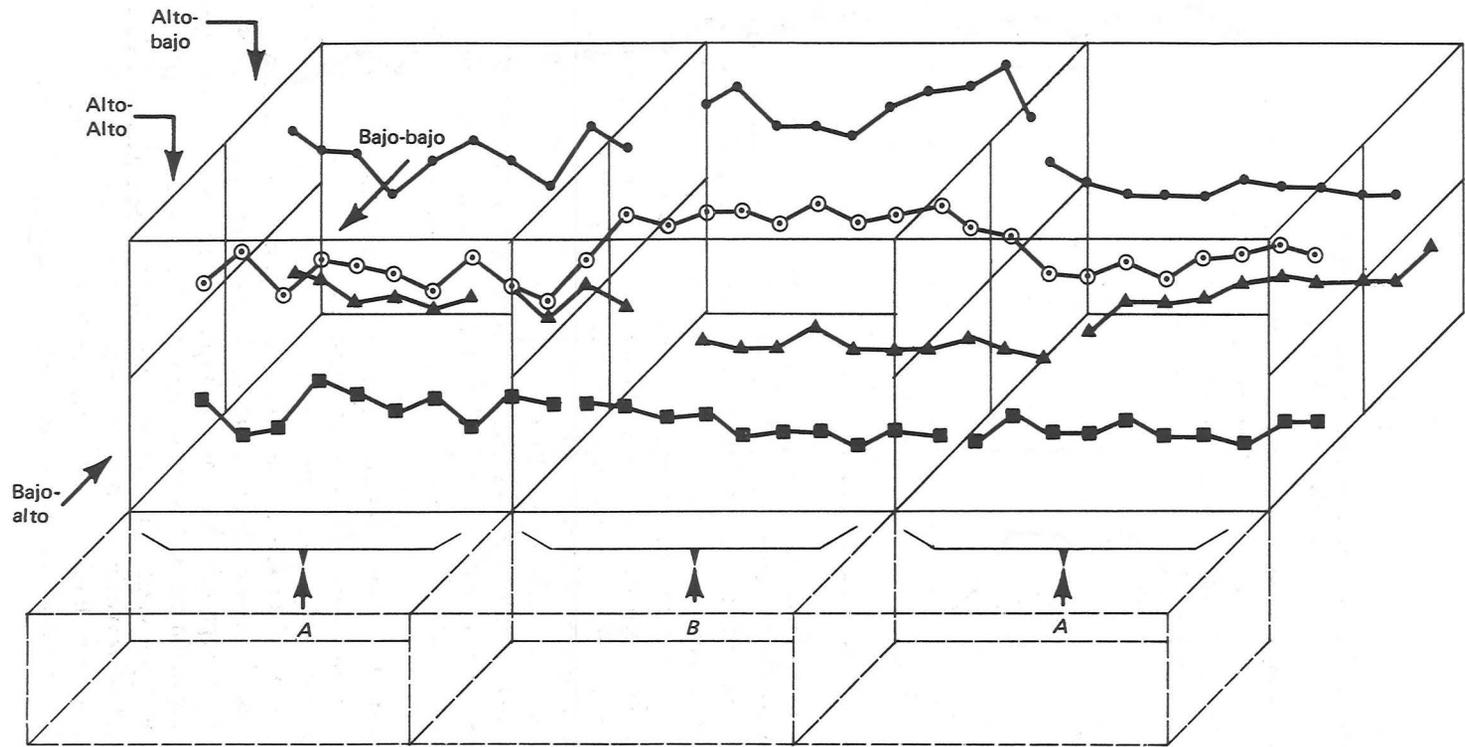
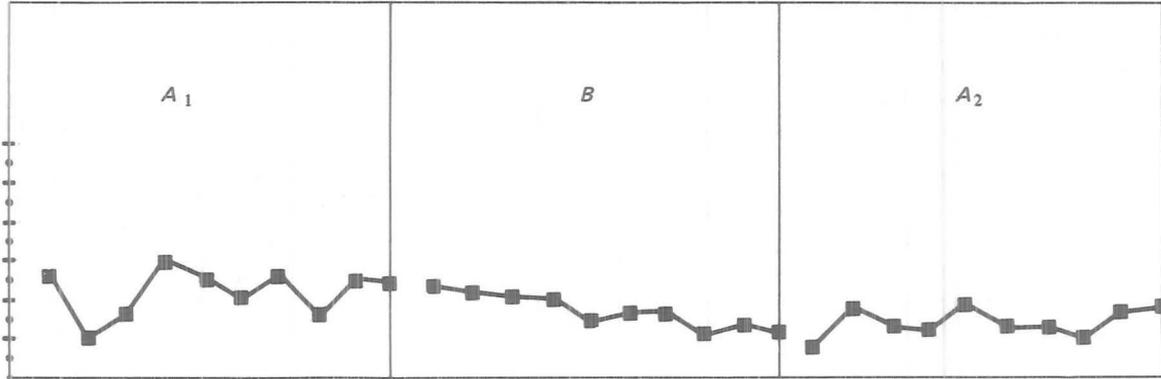
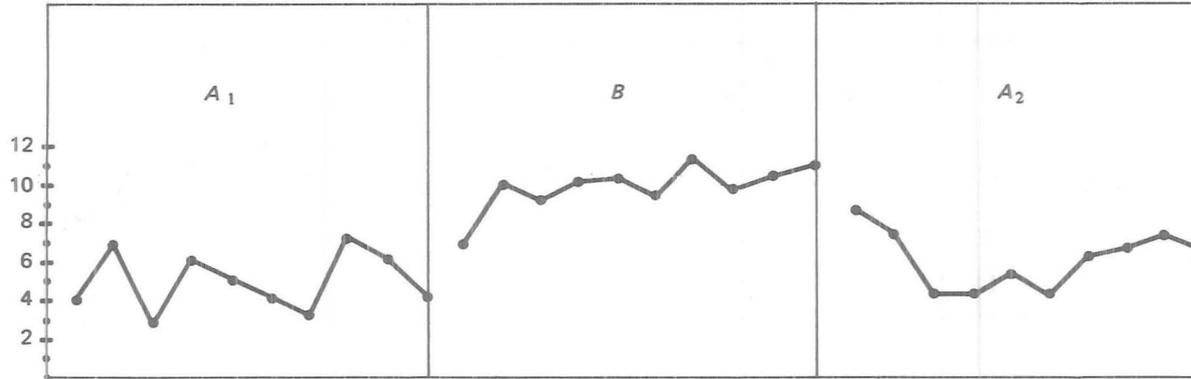
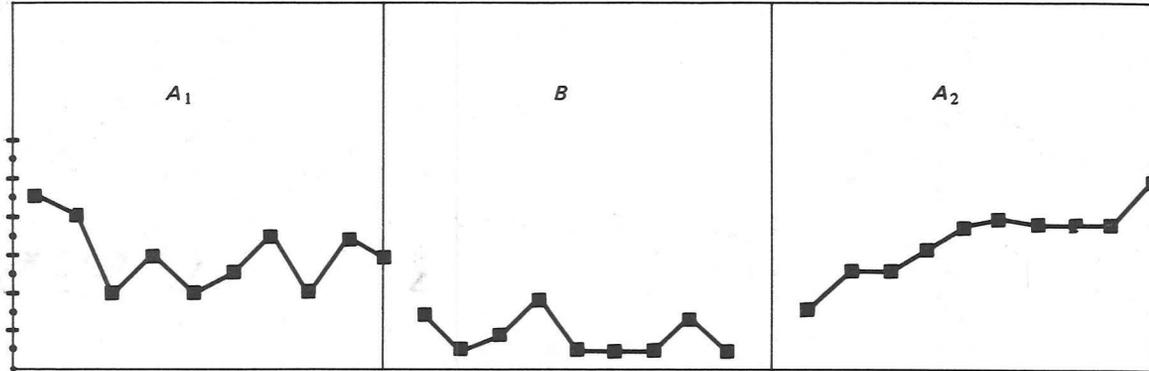
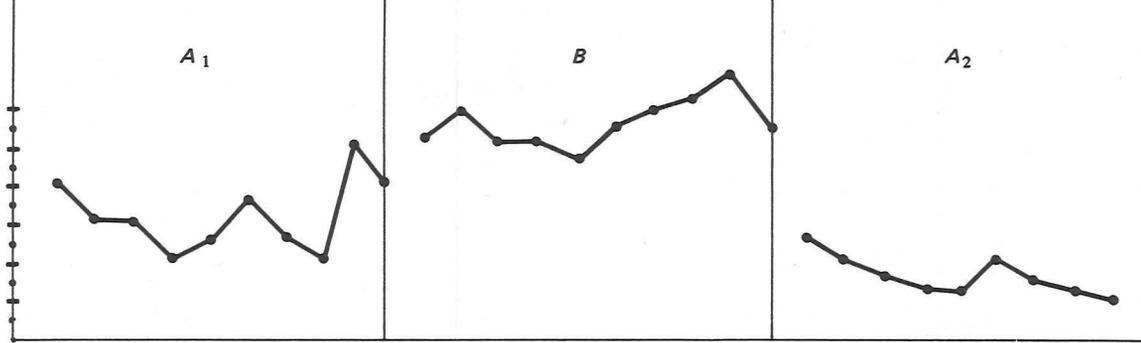


Figura 35. Representación tridimensional de datos ficticios obtenidos con un diseño combinado $2 \times 2 \times ABA$.

Alto-alto



Bajo-alto



Bajo-bajo

Figura 36. Representación bidimensional de datos ficticios de la figura 35 (diseño combinado $2 \times 2 \times ABA$).

Compárese este diseño con los diseños más simples que sirvieron para formarlo. Por un lado tenemos al diseño A-B-A; este diseño básicamente contestaría dos preguntas, una relativa al efecto que ejerce la VI sobre la línea base (fase manipulativa) y la otra referente a la reversibilidad o recuperación de la línea base original. Por otro lado tenemos al diseño factorial 2×2 , que es el otro contribuyente al diseño combinado. El diseño factorial 2×2 contesta básicamente tres preguntas: una, concerniente al efecto de la primera VI; otra, que se refiere al efecto de la segunda VI; y una tercera, relativa a su efecto conjunto.

El diseño combinado $2 \times 2 \times \text{ABA}$, por su parte, permitiría la apreciación de los efectos de cada una de las VI implicadas, de los efectos conjuntos de sus combinaciones y de una gran variedad de efectos locales. Estos efectos podrían valorarse tanto a través de la inspección visual de las gráficas como por medio de análisis estadísticos. Estos últimos son motivo de un estudio técnico que se puede tornar demasiado especializado y que queda por completo fuera del alcance de este libro. Así pues, las investigaciones acerca de este diseño en particular o de otras combinaciones potenciales más complejas en general, estarán dirigidas, primordialmente, a las preguntas que serán de legítimo interés experimental, minimizando sus implicaciones estadísticas. Una pregunta, para ser de legítimo interés experimental debe reunir, en este contexto, las siguientes características: *a*) debe implicar la comparación de por lo menos dos niveles operacionalmente diferentes (no necesariamente independientes); *b*) implicar información parcialmente redundante, pero no completamente redundante; *c*) debe ser relevante a los objetivos de una investigación particular. Por lo que respecta a la estadística, es factible la utilización de técnicas como el AVAR, a fin de analizar este tipo de diseño si se relajan ciertas restricciones y si se utiliza un modelo que describa adecuadamente al diseño. Como esto no es un hecho comprobado y debemos esperar los resultados de la investigación en esta área, el lector debe considerar la lógica de la estructura y la relevancia experimental de las preguntas que ésta puede ayudar a plantear y que se base en la inspección visual de los datos, hasta que los elementos del análisis estadístico estén disponibles (debidamente puestos a prueba y aprobados).

Las relaciones posibles en un diseño $2 \times 2 \times \text{ABA}$ parecen cumplir con los dos primeros criterios enunciados en el párrafo anterior, aunque no se pretende que las que aquí se mencionan constituyan una lista exhaustiva ni absolutamente correcta. Para facilitar la exposición, dividiremos a las VI en dos clases: “reversibles” y “factoriales”, a fin de denotar su procedencia de acuerdo a los diseños contribuyentes. La VI₃ “reversible” produce efectos sobre cada uno de los sujetos en

las cuatro diferentes celdillas del diseño factorial. Cuando se ignora la clasificación por la VI_2 "factorial", se aprecia el efecto "principal" de la VI_1 . Esto se lleva a cabo promediando los niveles "bajo" y "alto" de la VI_2 . Si se hace lo propio con la otra variable "factorial", se evalúa el efecto de la VI_2 . A su vez, es posible apreciar los efectos que ejercen estas dos VI sobre cada una de las fases (ABA) del diseño reversible.

Si se separa la secuencia A_1 -B del resto y se ignora la clasificación factorial, se puede notar el efecto global de la VI_3 . Si se separa ahora la secuencia B - A_2 , se podrá notar el efecto de la reversión. Finalmente, si se divide a estas secuencias en las cuatro condiciones factoriales, se podrán notar los efectos secundarios de la VI_3 y de la reversión. De acuerdo a este planteamiento, el diseño combinado podría producir datos acerca de un número de preguntas de investigación, parte de las cuales no aparecerían en un análisis tradicional de los datos. Dichas preguntas de investigación serían:

1. ¿Cuál es el efecto de la VI_1 "factorial" sobre la VD?
2. ¿Cuál es el efecto de la VI_2 "factorial" sobre la VD?
3. ¿Cuál es el efecto general de la VI_3 (a través de las tres fases e ignorando la clasificación factorial) "reversible" sobre la VD (línea base)?
4. ¿Cuál es el efecto de la aplicación de la VI_3 sobre la VD (comparación A_1 B)?
5. ¿Cuál es el efecto de la reversión de condiciones sobre la VD (comparación BA_2)?
6. ¿Qué tanto se logró la recuperación de la línea base original (comparación A_1A_2)?
7. ¿Cuál es el efecto conjunto de las dos VI "factoriales" ($VI_1 \times VI_2$) sobre la VD (independientemente de la clasificación ABA)?
8. ¿Cuál es el efecto de la VI_1 "factorial" sobre la VD a través de las fases A_1 B? (Nótese que durante la fase B habría un efecto conjunto de esta variable con la VI "reversible".)
9. ¿Cuál es el efecto de la VI_2 "factorial" sobre la VD durante las fases A_1 B? (También habría un efecto conjunto de esta variable con la VI "reversible" durante la fase B.)
10. ¿Cuál es el efecto de la VI_1 "factorial" sobre la VD a través de las fases BA_2 ?
11. ¿Cuál es el efecto de la VI_2 "factorial" sobre la VD durante las fases BA_2 ?
12. ¿Cuál es el efecto conjunto de la VI_1 y la VI_2 sobre la VD durante la fase A_1 ?

13. ¿Cuál es el efecto conjunto de la VI_1 y la VI_2 sobre la VD a través de la fase B? (Nótese que el efecto de la VI "reversible" sería concomitante al de estas dos variables "factoriales".)

14. ¿Cuál es el efecto conjunto de la VI_1 y la VI_2 sobre la VD durante la fase A_2 ?

15. ¿Cuál es el efecto conjunto de la VI_1 y la VI_2 sobre la VD a través de la secuencia A_1B ? (Nótese que durante la fase B las dos variables "factoriales" estarían interactuando con la variable "reversible".)

16. ¿Cuál es el efecto conjunto de la VI_1 y la VI_2 sobre la VD durante la secuencia BA_2 ? (También estaría presente la interacción con la variable "reversible" durante la fase B.)

17. ¿Cuál es el efecto conjunto de la VI_1 y la VI_2 sobre la VD a través de la secuencia ABA? (La misma interacción estaría presente durante la fase B.)

18. ¿Cuál es el efecto de la combinación de valores alto de la VI_1 , y alto de la VI_2 sobre la VD durante la secuencia A_1B ?

19. ¿Cuál es el efecto de la combinación de valores alto de la VI_1 y bajo de la VI_2 sobre la VD a través de la secuencia A_1B ?

20. ¿Cuál es el efecto de la combinación de valores bajo de la VI_1 y alto de la VI_2 sobre la VD durante la secuencia A_1B ?

21. ¿Cuál es el efecto de la combinación de valores bajo de la VI_1 y bajo de la VI_2 sobre la VD a través de la secuencia A_1B ?

22. ¿Cuál es el efecto de la combinación de valores alto de la VI_1 y alto de la VI_2 sobre la VD durante la secuencia BA_2 ?

23. ¿Cuál es el efecto de la combinación de valores alto de la VI_1 y bajo de la VI_2 sobre la VD a través de la secuencia BA_2 ?

24. ¿Cuál es el efecto de la combinación de valores bajo de la VI_1 y alto de la VI_2 sobre la VD durante la secuencia BA_2 ?

25. ¿Cuál es el efecto de la combinación de valores bajo de la VI_1 y bajo de la VI_2 sobre la VD a través de la secuencia BA_2 ?

26. ¿Cuál es el efecto de la combinación de valores alto de la VI_1 y alto de la VI_2 sobre la VD durante la secuencia ABA?

27. ¿Cuál es el efecto de la combinación de valores alto de la VI_1 y bajo de la VI_2 sobre la VD a través de la secuencia ABA?

28. ¿Cuál es el efecto de la combinación de valores bajo de la VI_1 y alto de la VI_2 sobre la VD durante la secuencia ABA?

29. ¿Cuál es el efecto de la combinación de valores bajo de la VI_1 y bajo de la VI_2 sobre la VD durante la secuencia ABA? (Nótese que las preguntas 18-29 corresponden a un corte a través de los pequeños cubos que se producen por la estructura del diseño factorial.)

Esta actividad de derivar preguntas de la estructura del diseño se puede continuar hasta incluir preguntas tan específicas (y posiblemente tan triviales) como: ¿Cuál es el efecto que ejerce la combinación de valores alto de la VI_1 y alto de la VI_2 sobre la VD durante la fase A_1 ?, etcétera. En general, se puede considerar que sólo las primeras 17 preguntas corresponden a lo que constituirían efectos principales e interacciones dobles y triples en diseños convencionales y que serían las únicas que debían formularse (y de éstas, las siete primeras se pueden contestar fácilmente con los diseños contribuyentes). Por otra parte, el resto de las preguntas pueden considerarse como correspondientes a efectos locales de importancia secundaria; pero en algunos casos podrían tener méritos suficientes como para que se les plantease. Cualquiera que fuese el número de preguntas que se pueden formular con este tipo de estructura (en este caso podrían exceder las cuarenta) e independientemente de su grado de importancia o trivialidad, parece evidente que un diseño combinado de este tipo puede ser definitivamente más informativo que sus diseños contribuyentes: un factorial de 2×2 con 120 observaciones o un ABA también con 120 observadores es, por tanto, más poderoso.

Este diseño incluye los detalles más característicos de sus diseños contribuyentes y, además, permite la evaluación de efectos específicos que aquéllos no podrían producir o advertir. Usando la noción fundamental del AVAR y contraviniendo en un número de casos la suposición de independencia estadística de las observaciones, debería ser posible apreciar estos efectos sutiles que el diseño permite producir. Obsérvese, por ejemplo, el cubo superior derecho a lo largo de las fases ABA de la figura 35. Un investigador podría estar interesado en la "interacción triple" $VI_1 \times VI_2 \times ABA$, la cual no sería posible apreciar con ninguno de los dos diseños contribuyentes por separado.

Una estructura como la del diseño $2 \times 2 \times ABA$ permite trabajar con un solo sujeto o con un grupo de sujetos por condición factorial. Las líneas que se ilustran en la figura 35 podrían representar las medias aritméticas de cada grupo. Como el efecto separado de la VI_3 "reversible" solamente puede ser estimado en esta estructura y no observado directamente, quizá algunos investigadores prefieran añadir un "sujeto control" con este propósito particular. Dicha opción está representada por el cubo extremo inferior en líneas quebradas en la figura 35 y no forma parte (necesariamente) del diseño principal.

Otra alternativa para obtener información basal sumamente importante que está ausente en el diseño $2 \times 2 \times ABA$, es proporcionada por un diseño $2 \times 2 \times ABCB$. En este último diseño se obtiene una línea base que es independiente de la estructura del diseño factorial.

Ésta aporta un nivel de comparación intrasujeto (o intragrupo) útil para la evaluación de los efectos de variables factoriales en las fases *B* (en las cuales no se incluye la variable reversible), así como los efectos conjuntos de la variable reversible y las variables factoriales. La línea base del diseño $2 \times 2 \times ABCB$ constituye el registro de la VD antes de la manipulación de las variables factoriales o de la variable reversible, y parece ser un nivel de comparación más confiable que el sujeto control mencionado anteriormente.

Por ejemplo, supóngase que un organismo se encontrará bajo el control de un programa de reforzamiento IV2, que las variables factoriales fueran la inclusión de un estímulo novedoso y un cambio en la proporción de intervalos grandes, y que la variable reversible fuera una inyección de pentobarbital sódico, la VD sería la tasa de respuesta.

Durante la etapa B_1 cada organismo de los cuatro requeridos por el diseño estaría sometido al programa IV2 y simultáneamente recibiría una combinación de dos valores de las variables factoriales (digamos, estímulo novedoso y una proporción de intervalos grandes de 0.20). Al pasar a la etapa *C* se incluye la variable reversible (el pentobarbital sódico). Finalmente, durante la fase B_2 se retira el pentobarbital y se vuelve a condiciones semejantes a las de la fase B_1 . Cualquier cambio que se observara a partir de la inclusión de la variable reversible nos indicaría el efecto de la interacción de dicha variable con las variables factoriales. Por otra parte, el comportamiento de la VD durante las fases B_1 y B_2 nos indicaría el efecto conjunto de las variables factoriales sobre la VD.

Sin embargo, a fin de poder apreciar adecuadamente el efecto de las VI (una reversible y dos factoriales) sería necesario contar con una línea base que sirviera como nivel de comparación. Esta es la función de la fase *A*; en este ejemplo esta línea base estaría constituida por el comportamiento de la VD bajo el programa VI2 antes de la inclusión de las variables factoriales o de la variable reversible.

Aunque diseños de este tipo lucen sumamente atractivos por su poder y aparente simplicidad en términos del trabajo implicado, también presentan desventajas que pudieran ser de cierta consideración. La primera de ellas se mencionó en párrafos anteriores y se relaciona con la necesidad de incluir sujetos control, a fin de observar directamente los efectos de determinada VI por sí sola. Otra desventaja obvia (que desde otro punto de vista se podría considerar como una ventaja) es la de que cada sujeto no recibe todas las combinaciones de valores de las VI (como sucedería en un diseño extendido A-B-C-A-D-A-E-A-F-A).

Por otra parte, presentan el gran atractivo de permitir una replicación de la relación encontrada en la secuencia ABA, condicionada a los efectos controlados de las variables "factoriales".

Si la proposición principal de que las estructuras de los diseños tradicionales son compatibles con las de los diseños conductuales es cierta, entonces sería posible desarrollar combinaciones que incluyeran, por ejemplo, a un diseño A-B de línea base múltiple por un lado y a un diseño multifactorial por otro. Podrían efectuarse combinaciones aún más complejas, las cuales, fundamentadas por datos, producirían implicaciones difíciles de seguir. Hay por lo menos dos consideraciones que deberían poner un límite superior al procedimiento de elaborar combinaciones cada vez más complejas. La primera de ellas se relaciona con el número de observaciones que se coleccionaría de tal manera. Este número podría ser enorme en algunos casos; se tendrían diseños sumamente costosos, cuyos análisis e interpretación no podrían hacerse sin la ayuda de una computadora y cuya ejecución sería monumental en términos del esfuerzo humano, equipo, materiales, etc.

La segunda aborda un aspecto fundamental: ¿De qué complejidad tendrían que ser las preguntas o series de preguntas a plantearse, a fin de justificar el uso de un diseño de tal magnitud? Parece obvio que, aun cuando se tiene al alcance la tecnología para el control y análisis de experimentos de este grado de complejidad, la ciencia de la conducta no ha producido los datos en los cuales se podrían cimentar complejos de preguntas apropiados para este tipo de diseño.

11.3. ANÁLISIS Y EVALUACIÓN DE LOS DATOS EN LAS CIENCIAS DE LA CONDUCTA

Como se ha visto repetidamente, el aspecto de la evaluación o análisis de los datos se puede convertir en un problema si las mediciones que se tienen no son sensibles, o están en error, o no son consistentes (confiables) o estables o son reactivas. La superación de cada una de estas desventajas simplemente proporciona un buen punto de partida para desarrollar análisis que sean útiles para la descripción, explicación y ulterior control de la conducta. Los análisis de los datos han cambiado de la aproximación discreta clásica basada en ensayos, errores (al responder), criterios arbitrarios de ejecución, etc., a la aproximación operante que hace descripciones completas de las condiciones en que ocurre la conducta y lleva a cabo registros continuos de aspectos específicos de ésta.

El énfasis en los procesos de transición (el "aprendizaje", por ejemplo) ha disminuido en favor de ciertos estados fijos, y los criterios de

ejecución ya no son impuestos arbitrariamente, sino que dependen de los parámetros de ejecución del sujeto particular y de la forma en que se manipulan las variables que controlan la conducta. Conforme se incrementa la complejidad de los problemas a resolver, los datos experimentales van requiriendo de análisis cada vez más finos que implican los medios tecnológicos disponibles. Los microanálisis contemporáneos dependen de la interacción entre el equipo de programación y el registro automático con computadoras que advierten cambios sumamente rápidos, tanto de la conducta como de sus controles. Estos desarrollos en el análisis de datos son tecnológicos y competen a la instrumentación; pero también tienen implicaciones metodológicas que incluyen al diseño experimental. Si, por ejemplo, el uso de técnicas análogas a la del "cuarto de vida", citada por Sidman (1960), se extendiese de tal manera que todo experimentador tuviera que valorar la calidad de su control antes de iniciar propiamente su experimento, entonces se debería incluir una etapa de calibración en la mayoría de los diseños experimentales.

Parece irónico que, a pesar de tener al alcance ciertas tecnologías eficaces para el análisis de datos conductuales, todavía no haya criterios uniformes para su evaluación. En el resto de esta sección se estudiará principalmente la valoración de alguna de las alternativas más factibles para este propósito.

Aunque el AVAR tradicional sería probablemente insensible para advertir ciertos cambios importantes en algunos diseños conductuales, especialmente de los tratados en la sección anterior, se puede utilizar su mecánica básica de sacar proporciones entre varianzas con efectividad si tan sólo se le enfoca desde un nuevo punto de vista. El modelo fijo del AVAR puede usarse apropiadamente para el análisis de datos de un sujeto o grupo de sujetos, sirviendo como su propio control. McNemar (1969) indicó que el uso repetido de los mismos sujetos bajo diferentes condiciones es preferible al empleo de distintos sujetos bajo diferentes condiciones. Esto no sería nada extraordinario si no proviniera de un destacado estadístico y las razones no fueran básicamente estadísticas. Además, señaló McNemar, la utilización de los mismos sujetos como su propio control da lugar a experimentos más económicos, mejor controlados y más precisos. La justificación es que las observaciones estarían altamente correlacionadas, lo cual produciría el efecto de reducir el error contenido en las diferencias entre las medias de las distintas condiciones.

Los dos componentes principales de la prueba F (el índice básico que produce el AVAR) apropiada en este caso son la *varianza (variabilidad) entre condiciones* y la *varianza (variabilidad) dentro de con-*

diciones. La primera varianza indica el monto de variabilidad producida alrededor de una media aritmética total (ideal) como resultado directo del efecto de las diferentes condiciones experimentales sobre la línea base. Esta varianza es análoga a la *varianza entre grupos*, usada en el análisis de diseños tradicionales de grupo. La segunda varianza indica el monto de variabilidad que existe dentro de cada condición. Con excepción de los efectos residuales de los diferentes tratamientos sobre las distintas etapas de la línea base, esta varianza se puede considerar como independiente de los tratamientos experimentales y generalmente refleja la falta de control experimental sobre la conducta del sujeto (de acuerdo con esta generalización, a mayor variabilidad, menor grado de control, aunque éste no es siempre el caso). El argumento lógico en el cual se basa la prueba F se puede formular como sigue: si la variabilidad producida por los tratamientos es de importancia científica, entonces deberá superar ampliamente la variabilidad "intrínseca" producida por deficiencias en el control experimental sobre la conducta del sujeto. Esta superioridad se puede estimar cuantitativamente mediante la simple computación de una proporción entre ambas varianzas: al dividir la *varianza entre condiciones* entre la *varianza dentro de las condiciones* se puede encontrar cuántas veces la primera supera a la segunda. Obviamente, si los efectos de los tratamientos experimentales son insignificantes, ambas varianzas se asemejarán y producirán un número que tiende a 1. Debe quedar claro que para interpretar el índice F obtenido, se debe tener en cuenta el número de condiciones experimentales en una línea base y también el número de observaciones dentro de cada condición. Ambos números constituyen los tradicionales "grados de libertad" y son los parámetros que sitúan a una determinada F en su lugar, dentro de una distribución hipotética de un número considerable de pruebas F . Tradicionalmente, se encuentra la probabilidad asociada a la ocurrencia de un valor F y se le interpreta como la probabilidad de que el valor obtenido (o uno más alto) se haya conseguido por la mera "acción del azar". Cuando dicha probabilidad es sumamente pequeña, se descarta la posibilidad de que el azar sea la causa del desbalanceo entre ambas varianzas (producto de la desigualdad entre las condiciones) y éste se atribuye a los tratamientos experimentales. En otras palabras, el efecto se infiere a través de la exclusión de la posibilidad de que el azar haya producido los datos obtenidos.

En una aproximación que no está basada en la inferencia estadística sino en el control experimental, el papel que desempeña una prueba como la F es el de ayudar a uniformizar los criterios de evaluación de los datos experimentales, buscándose así una mayor compatibilidad entre los diseños conductuales y los tradicionales.

En la presente aproximación al uso del AVAR, se propone que el investigador tiene la facultad de manipular por lo menos una de las dos varianzas implicadas en la proporción F y frecuentemente a ambas. Si esto es válido, el experimentador puede producir "efes" tan grandes o pequeñas como su control experimental sea efectivo. Una F con una probabilidad alta significa que el experimentador tiene un escaso control sobre la variabilidad implicada; una F con una probabilidad pequeña significa que tiene un efectivo control experimental. Nótese que el azar no desempeña ningún papel en este contexto.

Lo que un investigador tradicional conceptualizaría como "azar", desde este punto de vista se consideraría como correspondiente a alguna de estas deficiencias o combinaciones de ellas: *a*) errores de medición; *b*) variabilidad de muestreo, generalmente como producto de un número demasiado pequeño de observaciones; *c*) participación de VE como resultado de un escaso control experimental; *d*) efectos residuales activos o reactivos, y *e*) fuentes de invalidez interna. Quizá en este punto sería importante hacer una breve reinterpretación de los errores de tipo I y II en términos del control experimental.

Lo que en la aproximación clásica es un error de tipo I se puede deber, básicamente, a una de estas dos alternativas: *a*) a errores de medición, que producirían datos falsos que ocasionalmente conducirían al rechazo de la hipótesis nula (H_0), o *b*) a un débil control experimental, que permitiría que otras variables aparte de la VI produjeran datos espurios.

Lo que en la aproximación tradicional se considera como un error de tipo II, se puede conceptualizar como una deficiencia en la sensibilidad de las mediciones tomadas. Una medición insensible sería incapaz de reflejar diferencias reales entre dos varianzas (medias u otras estadísticas), apoyando falsamente a la H_0 .

Un experimentador riguroso no estaría dispuesto a que el azar fuera la causa de la variabilidad de sus datos, a menos que quisiera comparar esta variabilidad fuera de su control con la variabilidad bajo su control. Cuando el experimentador estuviese comparando varias condiciones de una línea base, trataría de aumentar la variabilidad entre las condiciones o fases de su diseño y de disminuir la variabilidad dentro de éstas. Esto lo haría por medio del control experimental. En este contexto su tarea sería la de producir índices F lo más altos posible, los cuales estarían asociados a probabilidades cada vez más pequeñas, disminuyendo así la participación de lo que constituye el llamado azar.

Cabe hacer notar que cuando no hay un "nivel de significancia" fijado de antemano, el investigador tiene una interpretación para todas y cada una de sus F s: todas las F s son "significativas". Una F baja

le indica que tiene que mejorar sus procedimientos o su instrumentación, y una F alta le indica cuál es la eficacia relativa de su control experimental. Además, una vez que se alcanza un nivel de probabilidad relativamente bajo (por ejemplo, 0.001), el investigador minucioso buscará un nuevo récord o nivel de probabilidad más exigente (por ejemplo, 0.0001 o 0.00001), en vez de contemplar con satisfacción el acostumbrado nivel de 0.05 o el de 0.01. Esta es la base del argumento expuesto por Meheel (1967) cuando comparó la aproximación de los físicos con la de los psicólogos tradicionales.

Dependiendo de la naturaleza de su problema y del nivel de probabilidad alcanzado, el investigador podrá planear procedimientos más eficaces o conformarse con el nivel alcanzado. Por otra parte, si la formulación estadística respecto a las características de la distribución F se aplica a este tipo de datos, entonces el tratar de incrementar la probabilidad más allá del 0.0001 (por ejemplo) no va a producir beneficios importantes, ya que hay pocos datos en la cola de la distribución, y una diferencia por un factor de 10 sería realmente insignificante.

A pesar de que en este empleo del AVAR se están contraviniendo varias de las suposiciones básicas de esta técnica, hay ciertas restricciones que se deben mantener si la prueba va a tener algún sentido. La primera es el número de observaciones por condición: cuando éste es insignificante, las variabilidades extremas pueden producir evaluaciones engañosas. Este riesgo es de dos sentidos: tanto pudiera indicarse falsamente un efecto donde no lo hay, como pasar desapercibido un verdadero efecto. (Ciertamente, se debe ser precavido al interpretar F s provenientes de menos de diez observaciones por condición.)

La segunda restricción obedece a la misma causa: cuando el número de observaciones por condición es insignificante, todas las condiciones deben contener aproximadamente la misma cantidad de observaciones. De no ser así, la prueba puede ser insensible para advertir auténticas diferencias.

Una tercera restricción, realmente irónica, es la siguiente: no se deben incluir líneas base perfectamente estables en este análisis. Cuando el control experimental es tan efectivo como para impedir que haya alguna variabilidad dentro de cada una de las condiciones (todos los datos son iguales y la varianza es ≈ 0), el AVAR no puede utilizarse, ya que esto implica por lo menos una división entre cero. Esto es irónico porque precisamente en esta situación el experimentador puede producir líneas base perfectas que cambian considerablemente de una condición a otra, donde el AVAR debería producir las F s más grandes; en vez de eso, es inservible en este caso. Por supuesto, basta con que exista una variabilidad mínima dentro de las condiciones, para que se

produzcan F s gigantescas bajo estas circunstancias (y, obviamente, dicha variabilidad podría ser fácilmente producida por el propio experimentador).

Otra técnica cuantitativa, que se propone por primera vez en este contexto, semejante al AVAR y que puede ser de utilidad para la evaluación de los diseños conductuales es la proporción de varianzas, la cual se puede considerar como una forma no ortodoxa de prueba F (señalada por Hays, 1963, en el contexto del diseño tradicional).

Mientras que el AVAR serviría para indicar si, en general, todas las condiciones del diseño fueron diferentes entre sí, el investigador podría preguntarse si la diferencia entre dos condiciones particulares fue importante o no. En este caso, una proporción de varianzas puede producir dicha información. Tomemos el caso de un diseño A-B, en el cual A es una fase meramente observacional y B es una fase manipulativa. En la primera, la varianza que se obtiene es observacional, $V(O)$, y en la segunda la varianza es impuesta, $V(I)$. La prueba F consistiría, simplemente, en la división $\frac{V(O)/n_1 - 1}{V(I)/n_2 - 1}$, donde n_1 es el número de observaciones en la primera condición y n_2 es el número de observaciones en la segunda condición. Las tres restricciones mencionadas con anterioridad son particularmente importantes en este caso, sobre todo, la segunda de ellas.

También puede darse el caso de que el experimentador, en forma deliberada, produzca una $V(I)$ extremadamente grande que daría lugar a una F pequeña. El problema de su interpretación ya lo han resuelto los estadísticos [véase, por ejemplo, Hays (1963) o Yamane (1970)] de una manera que gruesamente equivale a invertir la división y producir una F grande. (Técnicamente, la solución está dada encontrando los grados de libertad apropiados, buscar el valor de F en una tabla y sacar su recíproco $1/F$.)

Una tercera técnica para el análisis de datos conductuales es la empleada por Gott y Weiss (1972). Estos investigadores han usado el "modo armónico" para el análisis de los tiempos interrespuestas. Esta técnica puede extenderse al análisis de otros datos que se agrupan de manera "natural" alrededor de los "picos" de la gráfica (modos). De acuerdo a Gott y Weiss, las características principales de las armónicas modales son la presencia de picos o modos con amplitudes declinantes y la presencia de espacios intermodales de una magnitud aproximadamente constante. Desde luego, este tipo de análisis sólo se puede ejecutar sobre datos obtenidos con la mejor tecnología de control y medición.

Un índice que se propondrá en este libro para ayudar a determinar la estabilidad de los datos conductuales es la *proporción modal*. Bien se sabe que el modo es la medida estadística que mejor se adapta a los datos de la línea base conductual. Mientras otros índices, como la media aritmética o la mediana, son idealizaciones de un grupo de datos, el modo es una medida de frecuencia, o sea, es el valor que ocurre más a menudo en una serie de datos y corresponde necesariamente a un valor observado. La proporción modal es simplemente el número de veces que se repite el modo en relación al (dividido entre) número total de observaciones. Si tuviésemos una línea base con una estabilidad perfecta, todas las observaciones tendrían el mismo valor (por ejemplo, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6) y la proporción modal sería igual a 1.0. Ésta constituiría una forma extrema de variabilidad impuesta, la otra consistiría en crear la mayor variabilidad posible, en cuyo caso la proporción modal tendería a cero. Entonces, mientras más se acerque la proporción modal a 1, menor será la variabilidad y mayor la estabilidad de los datos. Algunas pruebas preliminares indican que la utilidad de esta medición es parcial hacia los casos de poca variabilidad; es decir, parece ser que el índice es insensible para discriminar el monto de la variabilidad para valores inferiores a 0.66.

Finalmente, se mencionó la estadística R_n , desarrollada por Revushky (1967) para estudiar datos que son el producto de la metodología del AEC. Esta técnica estadística requiere que el experimentador use un diseño de línea base múltiple, en el cual se le administra el tratamiento experimental a un organismo a la vez. La conducta de este organismo experimental se compara con la de los demás organismos que sirven como controles en esta comparación particular. Este proceso se repite hasta que todos los organismos han sido sometidos al tratamiento.

El orden en que los organismos pasan a ser sujetos experimentales se determina al azar. Revushky considera a cada aplicación del tratamiento como un subexperimento. En cada subexperimento se ordenan los sujetos por rango según su ejecución. La técnica R_n depende de que antes de la aplicación del tratamiento, cada organismo tenga las mismas probabilidades de ocupar cualquiera de los rangos posibles. El número de rangos va disminuyendo con cada subexperimento, siguiendo una serie $n, n-1, n-2, n-3, \dots, 1$.

Es necesario determinar cuáles son las probabilidades para cada subexperimento, y esto se consigue encontrando una *función generadora de probabilidad*. La estadística que se emplea para valorar la probabilidad de todo el experimento es la suma de los rangos resultantes en cada subexperimento. Obviamente, si todos los sujetos experimentales

obtuviesen consistentemente el rango más alto o el más bajo, se estarían desviando de lo que la función indicaría para rangos equiprobables. A fin de estimar la magnitud de esta desviación, Revushky elaboró tablas en las que indicó los niveles de significancia, el número de sujetos y el valor mínimo de R_n necesario para alcanzar determinado nivel de significancia. Asimismo, señaló que la técnica se puede extender para tratar diseños más complejos, pero que el investigador debería, en tales casos, desarrollar sus propias funciones generadoras de probabilidad.

Entre las principales características de la prueba (que implica tanto ventajas como desventajas), Revushky indicó: *a*) que el experimento se puede conducir sin ejecutar el último subexperimento; *b*) que, en vista de que el procedimiento requiere que los animales reciban el mismo procedimiento control repetidamente, habrá situaciones en las cuales esto no sea plausible o científicamente significativo; *c*) que cuando los efectos del tratamiento tardan en aparecer, las series de subexperimentos pueden ser excesivamente grandes (la técnica es más útil cuando el periodo de prueba es corto), y *d*) los parámetros de respuesta de diferentes animales pueden ser marcadamente distintos, en cuyo caso sería una pérdida de tiempo tratar de igualarlos.

Como se ha visto, no hay una incompatibilidad insalvable entre las formas de evaluación tradicionales y las conductuales. Parece ser que los factores que determinan la aparente incompatibilidad entre los análisis estadísticos y los datos conductuales son: por una parte, el planteamiento de una serie de restricciones irrealistas de parte de los modelos que fundamentan a los análisis, y por otra, la evitación sistemática de parte de múltiples investigadores entrenados con la metodología del AEC de la aplicación de las técnicas estadísticas disponibles.

Para terminar el estudio de este capítulo, es menester exponer una anotación acerca de los cambios que vamos a testificar en el diseño y análisis de experimentos conductuales. Al respecto, se puede decir que estamos asistiendo al rápido desarrollo de la ciencia de la conducta, y la base misma de ésta es el análisis experimental. Este último, hasta ahora relativamente indiferenciado y con un enfoque macroconductual, es probable que se divida en ramas tales como el análisis de estímulos, análisis de respuestas, análisis de reforzadores, y cuestiones similares. Es probable, también, que empiece a desarrollarse el análisis experimental a nivel microconductual, en el cual se estudiarían conductas tan minúsculas como los cambios instantáneos de los músculos faciales, respuestas tan rápidas y sutiles como la actividad cerebral implicada en la repetición verbal instantánea de estímulos previamente recibidos (la llamada "memoria") o la conducta de organismos cada

vez más pequeños (estableciéndose un nexo con la biología). Para el análisis de cambios tan diminutos o tan rápidos se requerirá no sólo un mejoramiento radical de la instrumentación existente, sino también un cambio en la actividad de "diseñar experimentos". Parece obvio que para estudiar continuamente la microconducta y para hacer los cambios pertinentes en las condiciones experimentales, se necesitarán diseños más dinámicos y flexibles con múltiples ramificaciones, cuyo uso estaría condicionado a los resultados de los tratamientos inmediatamente precedentes. Pocos experimentadores podrían mantener un paso acelerado de evaluación de resultados y monitoreo de las condiciones bajo circunstancias constantemente cambiantes. El análisis de los datos tendría que hacerse *en línea*, y las decisiones respecto a los cambios de procedimiento experimental tendrían que tomarse automáticamente por medio de computadoras. De esta manera, la labor del experimentador sería la de programar las estrategias más generales, indicando a la computadora los parámetros de decisión. El resto del trabajo sería automático.

11.4. LA ESTABILIDAD EN LOS DATOS CONDUCTUALES, CUANDO Y COMO SUSPENDER UNA LINEA BASE EN INVESTIGACION CON SUJETOS HUMANOS.

Los investigadores del comportamiento controlado experimentalmente, enfrentan a menudo el problema de decidir si sus datos se han estabilizado o no. Esto es particularmente cierto en laboratorios modestos, carentes de computadoras "en línea", para el análisis inmediato de los datos y, en muchas ocasiones, sin equipo de registro gráfico automático.

El problema de la estabilidad de los datos ya ha sido tratado con algún detalle (Sidman 1960, Boren 1966, Schoenfeld y Cole 1971) pero la mayoría de esos tratamientos se han restringido a respuestas operantes, particularmente a la tasa de respuestas. Esta aproximación ha dejado fuera de consideración otras formas de registro conductual que pueden ser una elaboración de la tasa de respuestas (por ejemplo, medidas temporales tales como distribución de tiempos interrespuesta o los cuartos de vida) o pueden ser medidas independientes de la tasa (tales como latencias, magnitudes, proporción de respuestas de apareamiento o número de respuestas correctas); además, la gran mayoría de los trabajos citados por dichos autores han utilizado sujetos no humanos durante periodos prolongados.

En esta sección se amplían las consideraciones acerca de la estabilidad para incluir una gran variedad de índices conductuales y evitar las confusiones que se pueden producir al generalizar a partir de lo que se ha escrito en relación con la estabilidad de la tasa de respuestas a otros índices de respuestas.

Por otra parte se enfocará la discusión al caso de la investigación con sujetos humanos por periodos relativamente muy cortos en comparación con la investigación animal. Este enfoque es propiciado por los particulares problemas que se producen cuando se trata de exponer a un sujeto humano a un tratamiento prolongado. Entre los principales problemas están la fatiga, el aburrimiento y la distractibilidad del mismo. Otro problema relacionado y, quizá el más importante, es el reducido grado de control experimental que se puede ejercer sobre este sujeto, en comparación con el de índole animal.

Como consecuencia de estas características peculiares del sujeto humano, en numerosas ocasiones se hace prácticamente imposible el desarrollo de un estado estable duradero. Por lo general, una vez que se ha adquirido un estado estable de corta duración, éste toma alguno de los siguientes cursos: *a*) la ejecución se deteriora, probablemente como una consecuencia de la fatiga, el aburrimiento o de algún distractor, o *b*) la ejecución mejora notablemente, al adaptarse el sujeto a las condiciones experimentales del tratamiento.

En ambos casos, el hecho de mantener las condiciones en busca de un estado estable más duradero puede producir resultados indeseables, tales como ejecuciones con gran variabilidad; por tanto, resultaría preferible tener un corto periodo de estabilidad, que buscar inútilmente una estabilidad más duradera.

¿Bajo qué criterios puede un investigador determinar si un índice conductual ha alcanzado una estabilidad limitada? Para contestar esta pregunta es necesario examinar los criterios de estabilidad que se plantean en casos más generales. Sidman (1960) señaló los siguientes:

- a*) Criterios derivados de estudios descriptivos.
- b*) Criterios de comparación de bloques de Schoenfeld y Cole.
- c*) Criterio de tiempo fijo o ajustado.
- d*) Criterios arbitrarios severos.
- e*) Criterios determinados por la variabilidad observada.
- f*) Criterios de inspección visual.

La mayoría de estos criterios involucran la manipulación y el registro sistemático de la conducta por periodos relativamente largos. Algunos [(*a*) y (*b*)] demandan de la ejecución un buen número de operaciones de cómputo, aunque no recurran a la estadística en su sentido más estrecho. Otros criterios [(*a*), (*e*) y (*f*)] requieren que el investigador esté bastante familiarizado con los valores: por ejemplo, el rango de variabilidad y la distribución característicos de sus variables de interés. El criterio de tiempo fijo puede dar como resultado una

descripción inadecuada del proceso conductual bajo estudio, al corresponder los datos a un estado de transición y no a uno fijo.

El criterio arbitrario severo puede plantear demandas inalcanzables en términos de costos experimentales (esfuerzo invertido en relación a la información obtenida). Con sujetos humanos se suma el problema de la exposición prolongada al control experimental.

(NOTA: La alternativa a esta posición sería el ejercicio del control experimental sobre los sujetos humanos por periodos prolongados tal y como se hace con otros organismos. Las implicaciones sociales y éticas de dicha aproximación son inconmensurables. Solamente en instituciones como el ejército y los hospitales psiquiátricos se podría intentar dicho control y su "justificación" sería altamente cuestionable.)

Sidman (1960), al discutir las características más generales de la selección de criterios, destacó el tipo de fenómeno investigado y el nivel de control experimental factible, como consideraciones fundamentales. En la investigación con sujetos humanos habría que añadir la fatiga, la habituación y "el aburrimiento"; también condicionó la calidad del criterio a la generación de relaciones funcionales replicables, ordenadas y significativas. La repetición de determinada función sería una forma de confirmación de la validez del criterio empleado. La forma más refinada de validación de un criterio sería la generalidad de los hallazgos.

Entre las observaciones y recomendaciones que este autor hace respecto a la selección y adopción de un criterio se encuentran:

- a) Al realizar experimentos con estados estables manipulados, el investigador puede excluir aspectos importantes del proceso bajo estudio.
- b) Lo inadecuado de un criterio puede evidenciarse cuando su experimento ya está muy avanzado; de cualquier modo, el diseño experimental debe prever esta eventualidad, permitiendo correcciones al criterio.

Por otra parte, una de las ventajas de sostener un criterio y apearse exclusivamente a él es evitar la selección de los datos que el investigador desea obtener de antemano.

Sidman considera el refinamiento de las técnicas de control como un requisito para la obtención de estados estables y, por tanto, de sensatos criterios de estabilidad.

La exposición de Sidman se extiende hasta incluir algunas formas de conducta aparentemente inestables, tales como las fluctuaciones cíclicas que pueden en sí mismas constituir un proceso conductual o las fluctuaciones derivadas de un precario control experimental. La descripción adecuada de las primeras (incluyendo, por ejemplo, el rango

de variabilidad, su relación con la hora del día, la duración de las pausas, etc.), permitiría la formulación de criterios de estabilidad para fluctuaciones cíclicas. En la investigación con sujetos humanos con limitada exposición temporal de éstos al control experimental, tales consideraciones parecen tener escasa aplicabilidad. El caso del control experimental precario no justifica la creación de un criterio de estabilidad y su recomendación se sintetiza en verificar todos los aspectos claves del procedimiento experimental, particularmente los relativos al equipo.

En ingeniería de control (Crownwen y cols. 1973), se considera la estabilidad como: "la habilidad de un sistema para restablecer un estado estable después de un disturbio de la entrada del sistema, en lugar de caer en una oscilación incontrolable . . . El sistema total debe ser suficientemente estable dentro del rango útil. La estabilidad de línea base es el mantenimiento de un valor constante sin corrimientos".

Es importante notar la analogía entre esta descripción y la hecha por Sidman, quien señala, en apoyo del criterio de inspección visual, la recuperación de la ejecución de línea base. Una demostración de esta clase de estabilidad se encuentra en el trabajo de Martínez (artículo en preparación) quien, en un estudio de discriminación-agresión, logró un gran número de recuperaciones del nivel de línea base, después de una disrupción bajo una variedad de condiciones experimentales. (Véase fig. 37).

Schoenfeld y Cole (1971) consideran que un sistema que tiende a la estabilidad pasará primero por un estado de transición; para decidir si se ha alcanzado la estabilidad, se plantean dos preguntas: ¿Ya pasó el estado de transición? y ¿Qué tan semejantes son dos mediciones sucesivas en cualquier fase del experimento? Señalan, además, dos fuentes de variabilidad que pueden condicionar la estabilidad: una es el producto de la transición hacia el "estado fijo final"; la otra depende de factores fuera del control experimental. Cuando el organismo se somete a un prolongado control experimental, la primera fuente de variabilidad pierde su importancia. En teoría es posible determinar el curso de la transición para separarlo de la variabilidad local, probablemente dependiente de los factores no controlados. En la práctica, señalan estos autores, es difícil hacer la separación, porque ésta depende parcialmente de la tasa de transición relativa de la tasa de muestreo de datos. Este problema se ve magnificado en la investigación con sujetos humanos, por lo menos en los siguientes respectos: si el estado de transición es lento y la muestra de observaciones, limitada, entonces el investigador puede observar un proceso aparentemente inestable y concluir que tiene una línea base no confiable; o peor aún, detectar

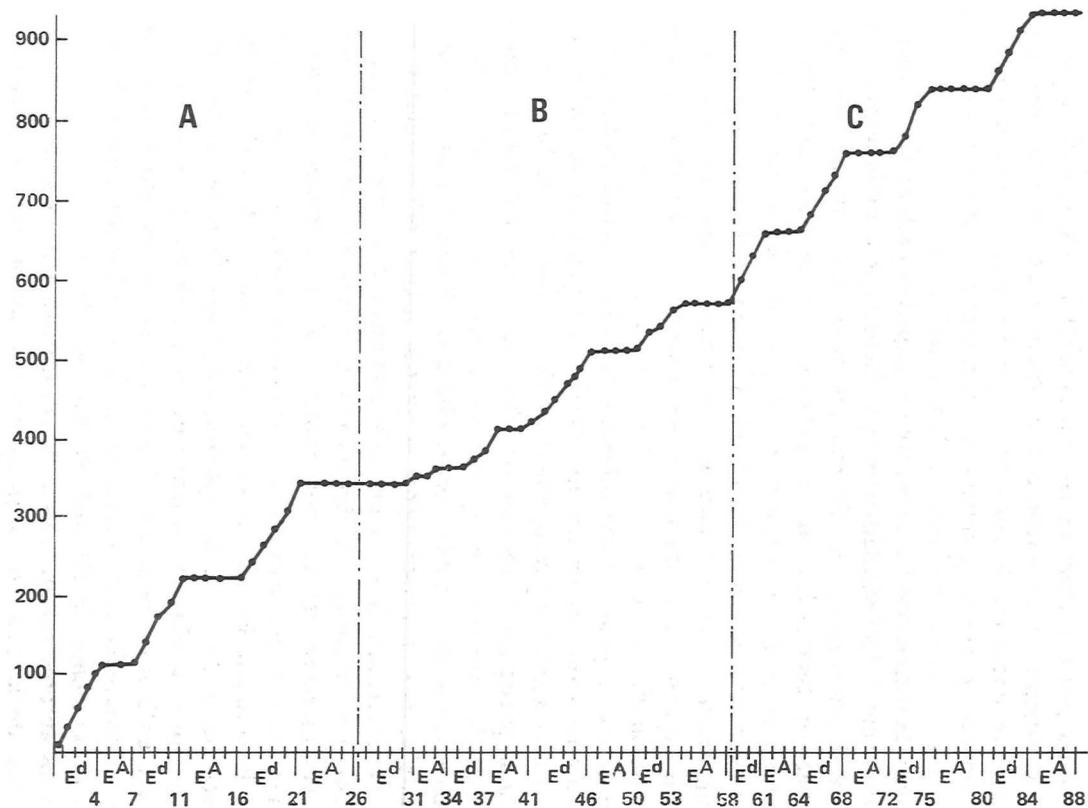


Figura 37. Datos de un estudio sobre discriminación-Agresión. Martínez (en preparación). La ejecución en las bases, está controlada por un programa concurrente: I V 2, R F 5. La fase B corresponde a la presentación visual de otro organismo.

sólo el principio de la transición y considerarla, erróneamente, como una muestra de conducta estable. Una alternativa para circunvenir este problema sería el hacer muestreos espaciados de bloques de observaciones; sin embargo, esto tendría dos dificultades principales: *a)* la pérdida de control experimental, que permitiría la intervención de una miríada de eventos en el intervalo entre bloques de observaciones que potencialmente alterarían los datos; *b)* que en cada bloque de mediciones sucesivas se podría reiniciar un proceso de transición que, de esta manera, no se completaría en forma definitiva.

Cuando la exposición a la situación experimental es muy corta, el problema de la variabilidad debida a factores no controlados o no detectados se puede sumar al de una transición incompleta; esto deja al investigador desprovisto de su principal herramienta para lograr una estabilidad confiable, el control prolongado sobre las variables que determinan la conducta de interés. Una salida de este callejón lo representa la utilización de criterios cuantitativos que, en ocasiones, relajan los requisitos de estabilidad determinada por el control experimental prolongado.

La segunda pregunta puede degenerar en una confusión fundamental. Si en una serie de mediciones sucesivas el valor de la última es afectado exclusivamente por la inmediatamente anterior, entonces estaríamos en la presencia de un proceso estocástico y, en tal caso, el conocimiento de la semejanza entre dos mediciones sucesivas sería claramente insuficiente para hablar de estabilidad. Todavía más: en una aproximación estocástica solamente se podría hablar de estabilidad en términos de matrices con uno o más estados absorbentes; o sea, se puede decir que un proceso ha alcanzado un estado fijo cuando la probabilidad de permanecer en él es igual a 1.00. Para determinar si un proceso estocástico ha entrado en un estado absorbente se debería observar indefinidamente la respuesta característica de dicho estado (al menos en teoría). En la práctica, esta restricción se ha relajado para dar lugar a criterios arbitrarios. Claramente, la pregunta de Schoenfeld y Cole no tenía la connotación de un proceso estocástico, sino más bien indicaba la comparación entre series adyacentes de observaciones, lo cual se detecta en su criterio.

Al considerar su propio criterio de comparación de bloques, Schoenfeld y Cole (1971) encontraron que, en ocasiones, su cumplimiento no implica la obtención de un estado estable "final". El criterio se puede alcanzar y, sin embargo, la ejecución puede ser atípica en relación a sesiones posteriores o a la tendencia total del experimento. Sidley y Schoenfeld (1964) habían resuelto este problema al determinar que con 40 o 50 días de exposición a las condiciones experimentales se

abarcarían los valores típicos de exposiciones más prolongadas. Algunas de las incoherencias encontradas con los criterios de estabilidad dependientes de las respuestas han favorecido, no obstante, a los criterios basados en duraciones de exposición a las condiciones experimentales preespecificadas. En aquella investigación humana en la cual esta exposición debe ser limitada puede ocurrir una rápida transición a un estado "final" relativamente estable, seguida de un deterioro en el nivel de ejecución, como ocurrió en estudios de respuesta retardada (Sierra y Castro, manuscrito inédito) o dado un número predeterminado de ensayos, éste puede resultar insuficiente para apreciar la evolución a un estado manipulado, estable o semiestable. La recomendación de Sidman, relativa a desarrollar estudios descriptivos para determinar la duración óptima de un criterio de estabilidad, cobra mayor importancia al considerar la investigación con sujetos humanos, en la cual la exposición a los tratamientos experimentales es limitada.

En un artículo reciente, Blough (1975) remarcó la importancia de alcanzar un nivel asintótico para la evaluación de gradientes de generalización operante; tales asíntotas fueron cruciales para determinar si un gradiente de control de estímulos es positivo o negativo. Por otra parte, Boren (1966) consideró el problema de la estabilidad de la línea base en relación a los efectos de las drogas sobre la conducta y, a pesar de estar de acuerdo en la deseabilidad de los estados estables en general, hizo notar algunas excepciones importantes; ellas son:

- a) Mientras una estabilidad extrema puede demostrar un alto grado de control experimental, puede también requerir una dosis masiva de una droga para alterar la estabilidad, lo que pudiera tener consecuencias adversas, tales como la de hacer atáxica a una rata o postrarla gravemente.
- b) En estudios con drogas se pueden encontrar casos en que la conducta es desviada o se adapta a las contingencias comunes;; estas conductas serían generalmente inestables (quizá sufran cambios lentos y sistemáticos) y, sin embargo, interesantes y factibles de tratarse.
- c) Hay conductas semiestables mantenidas por contingencias pasadas o espurias.

Cuando el sujeto humano puede ser observado bajo condiciones controladas durante periodos prolongados, es factible aplicar las mismas normas de control de la variabilidad ya examinadas y resolver así el problema de la estabilidad. Cuando esto no es posible (y en este trabajo se supone que la mayoría de la investigación con sujetos humanos está en ese caso) el investigador puede tomar cualquiera de los siguientes cursos de acción:

a) *Aplicar criterios arbitrarios, de exposición del sujeto a condiciones de línea base.* Si estos criterios son demasiado cortos pueden producir datos con una variabilidad intolerablemente grande; pueden también representar un estado de transición incompleto; el usar estos datos como nivel de comparación puede confundir la interpretación y evaluación de los tratamientos experimentales. Por otra parte, si un criterio es demasiado tardado, entonces puede volverse prácticamente inalcanzable con consecuencias indeseables, tales como una gran mortalidad experimental o una variabilidad. También caben dentro de esta categoría los criterios usados en la investigación de conducta conceptual humana, en los cuales se determina arbitrariamente si un sujeto ha “aprendido” un determinado concepto al responder “correctamente” un número (arbitrario) sucesivo de ensayos. Qué tan representativo de un proceso es un criterio de este tipo, permanece abierto a la discusión; en síntesis, estos criterios representan más desventajas que ventajas aun cuando son preferibles a los llamados “pre-tests”, en cuanto aportan más de una medición y —en ocasiones— pueden permitir la observación de ejecuciones bastante estables. Siempre que resulte posible sería preferible optar por cualquiera de las siguientes alternativas:

b) *Para cada línea de investigación, llevar a cabo estudios descriptivos que indiquen el rango de variabilidad aceptable y el tiempo mínimo deseable, de condición basal.* Obviamente, ésta es la posición más parsimoniosa y aplicable a cualquier tipo de investigación experimental; por otro lado, demanda (en cuanto condiciona los estudios manipulados) mayor interés al desarrollo de estudios descriptivos no tan interesantes. Una serie de estudios en la cual se varíe el tiempo o el número de ensayos bajo condiciones basales, podría indicar, por lo menos en forma gruesa, cuál es la variabilidad esperable y así, determinar el máximo soportable. También se puede detectar cuál es la exposición mínima requerida y, en ocasiones, cuál el punto donde la ejecución realizada por los sujetos empieza a deteriorarse notoriamente. Llevar a los sujetos a este extremo es la forma más cercana a la del control experimental prolongado, que se logra con otros organismos.

c) *Usar uno o más estados semiestables enfatizando la tendencia de los datos.* Hay numerosos ejemplos de datos en la investigación con sujetos humanos que se aproximan a un estado estable sin alcanzar en verdad un criterio más o menos estricto, como el de Schoenfeld y Cole; en tales casos, es frecuente encontrar “corridas” cortas de seis o siete puntos, en los que un índice conductual señala el mismo valor (o valores muy semejantes) para después desviarse de dicho nivel. Esto puede suceder a lo largo de una sesión o fase experimental. Los datos de esta clase pueden considerarse como semiestables y pueden utilizarse,

con alguna confianza, como nivel de comparación. En vista de que los parámetros más populares como la media o el modo resultan poco informativos (y aun engañosos), en estas circunstancias, el nivel general de los datos y, particularmente su tendencia, pueden resultar mejores índices de un estado semiestable.

Un ejemplo de dicha situación se presenta en la figura 38, y Urbina y Col. (manuscrito inédito) en un estudio de respuesta retardada analizaron sus datos por ensayos y por posición serial. Estos datos (por ensayos) pueden caracterizarse por medio de la ecuación de la línea recta $Y = a + bX$ donde el primer parámetro (a) se puede interpretar como el nivel de la recta, y el segundo (b) como la tendencia. En estos datos las diferencias que se evidencian visualmente se cuantifican en los parámetros de las ecuaciones 1 y 2.

$$Y = 0.5994 + (-0.0013)20.5 \quad (1)$$

$$Y = 0.3250 + 0.0059(7) \quad (2)$$

Donde la primera ecuación describe los datos de línea base mientras la segunda lo hace con los datos de la fase experimental. Los datos de línea base fueron divididos en bloques en 10 ensayos cada uno tomándose el bloque más estable y fueron transformados a una curva de posición serial. La comparación entre esta curva y la curva de posición serial derivada de la sesión completa con 40 ensayos se presenta en la fig. 39. Es clara la semejanza entre ambas curvas. Este fue el caso para la gran mayoría (90%) de las comparaciones de este tipo, realizadas en dicho estudio. Los otros dos casos que ocurrieron, fueron: *a*) una diferencia notable entre ambas curvas (representada en la figura 40), y *b*) una representación más aguda de las características de la curva, por otra parte de los datos estables (representada en la figura 41). Desde el punto de vista de la estadística es técnicamente correcto hacer comparaciones entre las pendientes (tendencias), sin que la restricción respecto a la independencia de las observaciones tenga importancia.

d) Desarrollar criterios cuantitativos. El desarrollo de criterios cuantitativos no implica necesariamente la aplicación de la estadística tradicional. Los criterios de Schoenfeld y Cole, Castro, Revushky, Gott y Weis (sección 11-3), ejemplifican esta situación.

Un criterio más (Castro, datos inéditos) se basa en la teoría de la medida: según está el "puntaje verdadero" de un sujeto, se puede aproximar cuando un número de observaciones tiende a infinito. En una forma más realista, el mismo argumento es válido para un número

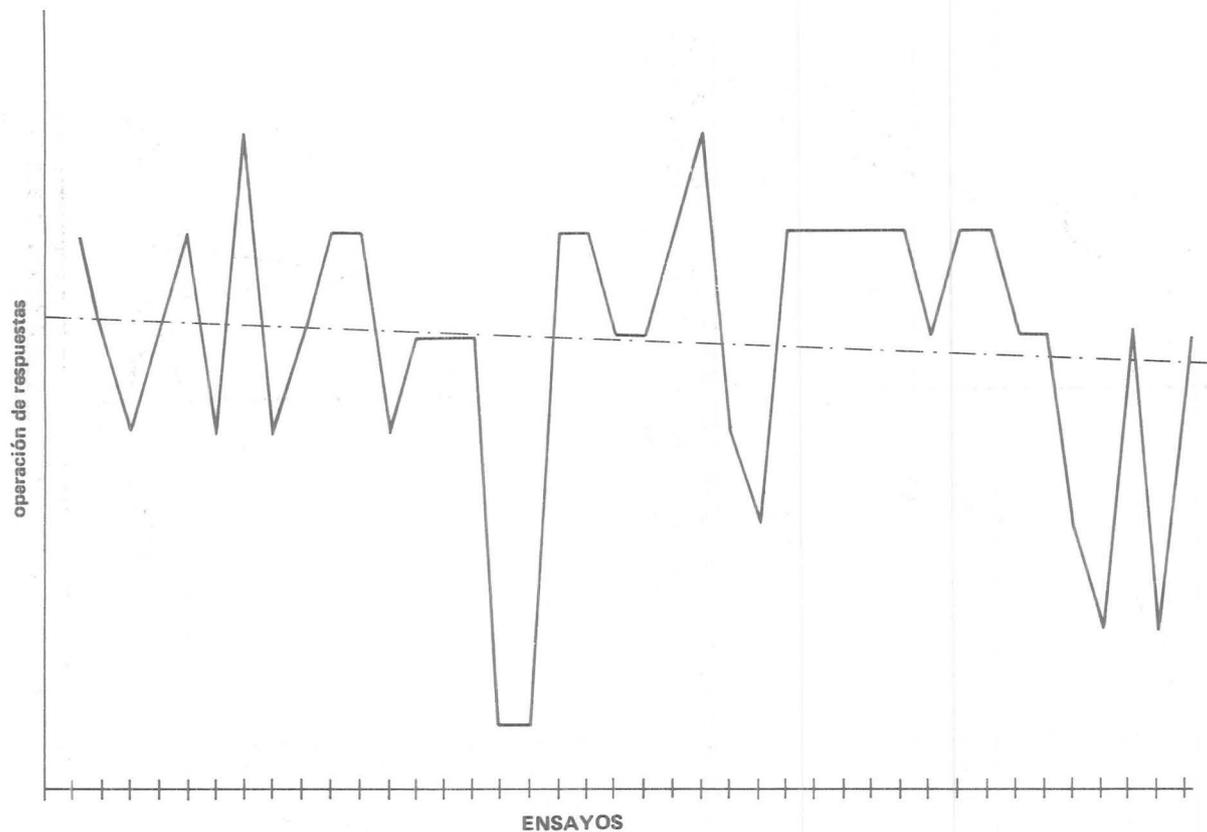


Figura 38. Datos de línea base, obtenidos de la serie completa de ensayos (40 en total).

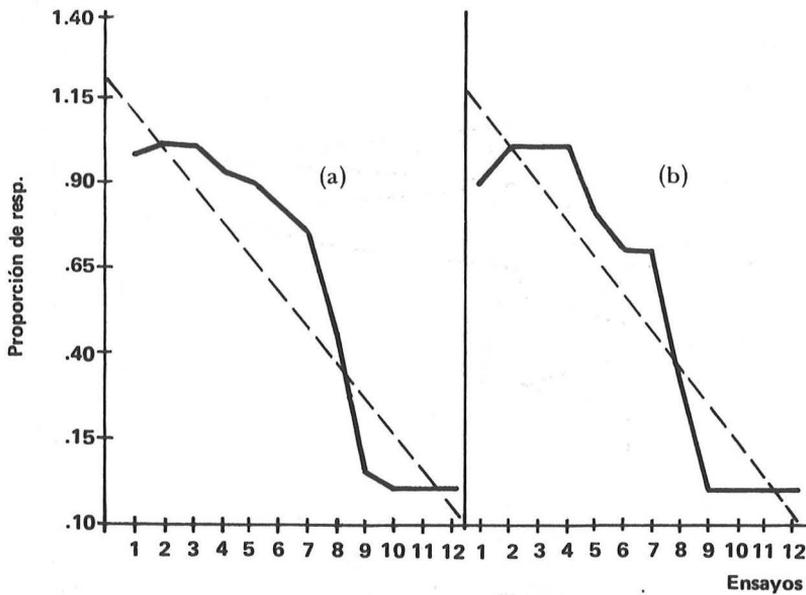


Figura 39 (a). Curva de posición serial obtenida de la serie completa de observaciones (40 observaciones) (b). Curva de posición serial obtenida exclusivamente con el bloque de observaciones más estables (10 observaciones)

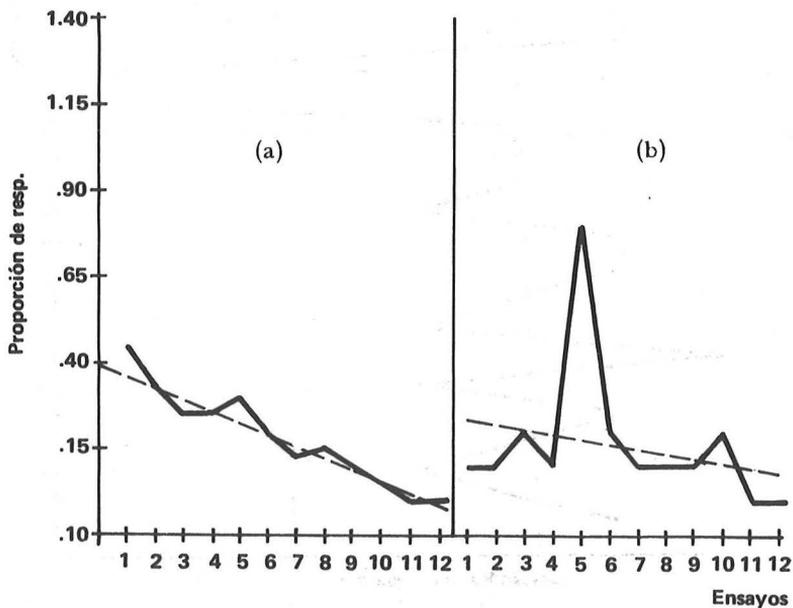


Figura 40 (a). Curva de posición serial obtenida con la serie completa de observaciones (40 observaciones) (b). Curva de posición serial obtenida exclusivamente con el bloque de observaciones más estables (10 observaciones).

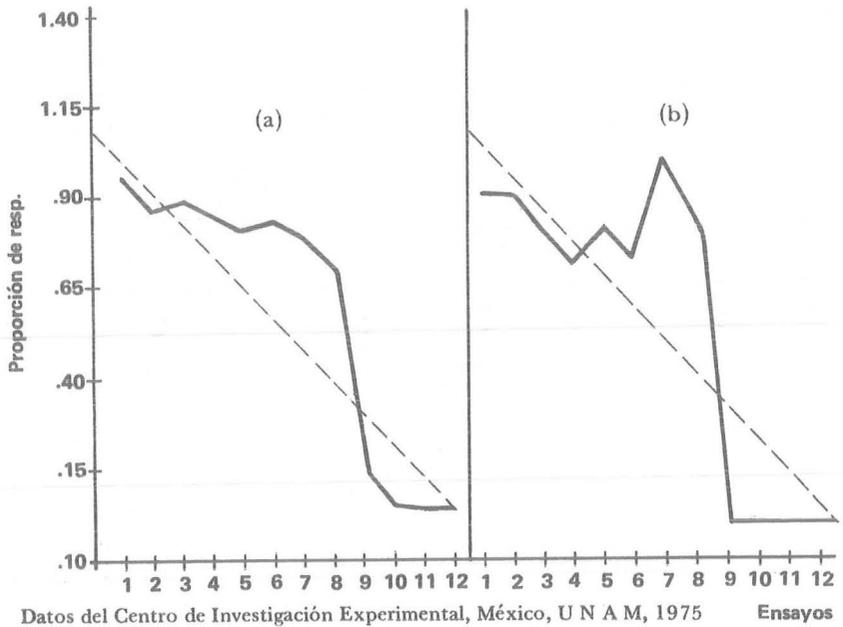


Figura 41 (a). Curva de posición serial obtenida con la serie completa de observaciones (40 observaciones) (b). Curva de posición serial obtenida exclusivamente con el bloque de observaciones más estables (10 observaciones).

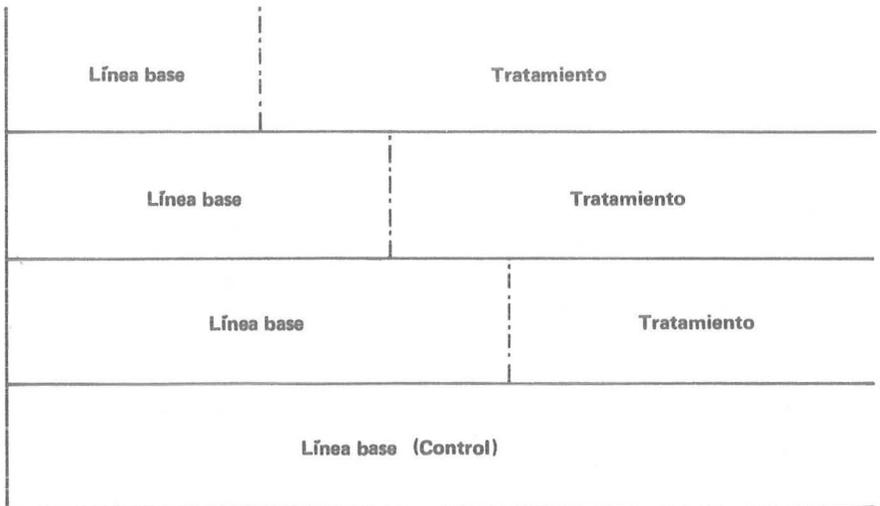


Figura 42. Diagrama del diseño experimental mixto

infinito de observaciones, de modo que la media aritmética se incluye a la última observación de una serie que se aproximara más al puntaje real que ninguna otra media que contenga menos observaciones; en esta técnica se computan todas las medias sucesivas de una serie de observaciones, se restan de la media total, se elevan al cuadrado y se suman, obteniéndose así una suma de cuadrados. Las observaciones se agrupan en bloques y se elige como más estable el nivel que arroja la menor suma de cuadrados. Aunque este procedimiento tiende a favorecer al último bloque, no siempre es el que resulta menor.

Todos estos criterios deberían tender a determinar un punto de corte de la línea base, a partir del cual la probabilidad de que una medida determinada caiga dentro de un rango de variabilidad estrecho es bastante alta.

e) Usar diseños experimentales mixtos. Cuando se prevé que el nivel de referencia que proporciona una línea base no será lo suficientemente confiable, se pueden organizar las observaciones de interés en la forma de un diseño experimental mixto; es decir, aquél en que se combinan la línea base y su grupo control (Castro, 1975). Un ejemplo de este tipo de diseño se presenta en la figura 42.

Esta figura representa un diseño de línea base múltiple en la cual la última condición es siempre de línea base, por lo cual funge como condición control y sirve para que los tratamientos puedan evaluarse tanto contra sus propios controles como contra este control independiente.

Ciertamente el problema de estabilidad de los datos conductuales es, en esencia, de control experimental; sin embargo, cuando éste no resulta asequible en el grado deseable, pasa a ser, en verdad, un problema de medición o de diseño.

De las alternativas consideradas, la más parsimoniosa es la propuesta por Sidman, relativa al desarrollo de estudios descriptivos; por ello, debiera seguirse siempre que fuese posible. Cuando esto no es así, el uso de estados semiestables que analizan la tendencia y nivel de los datos parece ser la alternativa más promisoría.

Por otra parte, la investigación de criterios de estabilidad está lejos de considerarse como completa y las posibilidades mencionadas son solamente una muestra de lo que los investigadores podrían hacer en esta área.

referencias bibliográficas

- Anderson, N. H. *Scales and statistics: parametric and nonparametric in statistical issues: A reader for the behavioral sciences*, Wadsworth Publishing Co., 1972.
- Atkinson, R. C., Bowers, G. H. y Crothers, E. J. *An introduction to mathematical learning theory*, John Wiley & Sons, Inc., 1965.
- Azrin, N., Robin, H., O'Brien, F., Ayllon, T. y Roll, O. Behavioral engineering: postural control by a portable operant apparatus, *Journal of Applied Behavior Analysis* vol. I, núm. 2, 1968, págs. 99-108.
- Bachrach, A. J. *Psychological research: an introduction*, Random House, Inc., 1972.
- Baer, D. M., Wolf, M. M. y Risley, T. R. Some current dimensions of applied behavior analysis, *JABA*, 1968, 1, págs. 91-97.
- Bailey, J. S., Wolf, M. M. y Phillips, E. L. *Home-based reinforcement and the modification of pre-delinquents' classroom behavior*, *JABA*, 1970, 3, págs. 223-233.
- Baan, D. *The test of significance in psychological research*. *Psychological Bulletin*. 1966, 66, págs. 423-437.
- Baker, F. B. *Experimental design consideration associated with large-scale research projects in improving experimental design and statistical analysis*, Stanley J. Rand McNally & Co., 1967.
- Baker, R. The use of operant conditioning to reinstate speech in mute schizophrenics, *Behavior Research and Therapy*, 1971, 9, págs. 329-336.
- Bandura, A. *Principles of behavior modification*, Holt, Rinehart & Winston, 1969.
- Barber, P. y Folkard, S. Reaction time under stimulus uncertainty with response certainty, *Journal of Experimental Psychology*, 1972, vol. 93, número 1, págs. 138-142.
- Barton, Elizabeth S., Guess, D., García, E. y Baer, D. M. Improvement of retardates meantime behaviors by time ont procedures using multiple baseline techniques, *JABA*, 1970, 3, págs. 77-84.
- Bensberg, G. J., Colwell, C. N. y Cassel, R. H. Teaching the profoundly retarded self-help activities by behavior shaping techniques, *American Journal of Mental Deficiency*, 1965, 69, págs. 674-679.
- Blough, D. S. Steady State Data and a quantitative model of operant generalization and discrimination, *Journal of Experimental Psychology: Animal Behavior Processes*, 1975. Vol. 104, núm. 1, 3-21.

- Boreau, C. A. *A note on measurement scales and statistical tests in statistical issues: A reader for the behavioral sciences*, Wadsworth Publishing Co. 1972.
- Boren, J. J. *The study of drugs with operant techniques in operant behavior: areas of research and application*. Honig, W. K. 1966. Appleton-Century-Crofts.
- Bourne, L. E. y Miller, S. Some stimulus variables affecting solution shift performance. *J. Exp.* 4, 1973, vol. 98, núm. 2, págs. 291-296.
- Breyer, N. L. y Axelron, S. *Managing behavior, part 6: summaries of selected behavior modification studies*. H. & H. Enterprises, Inc. 1971.
- Briggs, G. E. y Swanson, J. M. Encoding, decoding and central function in human information processing. *Journal of Experimental Psychology*. 1970, vol. 86, núm. 2, págs. 296-308.
- Brodén, M., Hall, R. V. y Mitts, B. *The effect of self-recording on the classroom behavior of two eighth-grade students*, JABA, 1971, 4, págs. 191-199.
- Brown, L. y Foshee, J. C. Comparative techniques for increasing attending behavior of retarded students, *Education and Training of the Mentally Retarded*, 1971, 6, págs. 4-11.
- Burchard, J. D. y Barrera, F. An analysis of Timeont and response cost in a programmed environment, JABA, 1972, 5, págs. 271-282.
- Calkins, D. S. An empirical investigation of the effects of the violation of the assumption that the covariable in analysis of covariance is a fixed variable, Tesis doctoral inédita, University of Texas, Austin, 1971.
- Campbell, D. T. *Administrative experimentation institutional records, and nonreactive measures in improving experimental design and statistical analysis*. Stanley, J. (dir.), Rand McNally, 1967.
- Campbell, D. T. y Stanley, J. C. *Experimental and quasi-experimental designs for research*, Rand McNally, 1966.
- Campione, J. C. y Beaton, V. L. Transfer of training: some boundary conditions and initial theory. *Journal of Experimental Child Psychology*. 13, págs. 94-114, 1972.
- Cañedo, A. (Comunicación personal), 1972.
- Castro, L. (Datos inéditos), 1975 b.
- Castro, L. y Escabí, H. (Datos inéditos), 1973.
- Cemek, L. S., Sagostsky, G. y Moshier, C. Development of the ability to encode within evaluative dimensions. *Journal of Experimental Child Psychology*, 13, págs. 210-219, 1972.
- Clement, P. W., Fazzone, R. A. y Goldstein, B. Tangible reinforcers and child group therapy. *American Academy of Child Psychiatry Journal*. 1970, 7, págs. 409-427.
- Cochran, W. G. y Cox, G. M. *Diseños experimentales*, Editorial Trillas, México, 1971.
- Collins, D., Kessen, W. y Haith, M. Note on an attempt to replicate a relation between stimulus unpredictability and infant attention. *Journal of Experimental Child Psychology*, 13, págs. 1-8, 1972.
- Cooper, Margaret L., Tomson, Carolyn L. y Baer, D. M. *The experimental modification of teacher attending behavior*, JABA, 1970, 3, págs. 153-157.

- Craig, H. y Holland, A. Reinforcement of visual attending in classrooms for deaf children, *Journal of Applied Behavior Analysis*. 1970, 3, págs. 97-109.
- Crowell, L. Weibell, F. J., Pfeiffer, E. A. y Usselman, L. B. *Biomedical Instrumentation and Measurement*, Prentice Hall, 1973.
- Cherek, D. R., Thompson, T. y Heistad, G. T. *Responding maintained by the opportunity to attack during an internal food reinforcement schedule*. JEAB, 1973, 19, págs. 113-123.
- Christopherson, E. R., Arnold, C. M., Hill, D. W. y Quilitich, H. R. *The home point system: token reinforcement procedures for application by parents of children with behavior problems*, JABA, 1972, 5, págs. 485-497.
- D'Amato, M. R. *Experimental psychology: methodology, psychophysics and learning*, McGraw-Hill, 1970.
- DeVilliers, P. A. y Millenson, J. R. Concurrent performances: a baseline for the study of conditioned anxiety. *Journal of Experimental Analysis of Behavior*. 1972, 18, págs. 287-294.
- Dingman, H. F. *Scientific method and reproducibility of results*. Manuscrito inédito, 1969.
- Dixon, W. J. y Massey, F. J. *Introduction to statistical analysis*, McGraw-Hill, Inc. 1969.
- Downey, J. y Harrison, J. M. Control of responding by location of auditory stimule: role of differential and non-differential reinforcement, JEAB, 1972, 18, págs. 453-462.
- Dudycha, A. L. y Dudycha, L. W. *Behavioral statistics: an historical perspective in statistical issues: a reader for the behavioral sciences*. Kirk, R. E. (dir.), Wadsworth Publishing Co., 1972.
- Dukes, W. F. M. = 1. *Psychological Bulletin*, 1965, 64, págs. 74-79.
- Dunham, P. J. Some effects of punishment upon unpunished responding. *Journal of the Experimental Analysis of Behavior*, 1972, 17, págs. 443-450.
- Edgington, E. G. Randomization tests for one-subject operant experiments, *The Journal of Psychology*, 1975, 90, 57-68.
- Edwards, A. L. *Experimental design in psychological research*, Holt, Rinehart & Winston, Inc. 1968.
- Edwards, W. Tactical note on the relation between scientific and statistical hypotheses in statistical issues: A reader for the behavioral sciences, Kirk, R. E. Brooks/Cole Publishing Co. 1972.
- Evans, R. I. B. F. *Skinner: The man and his ideas*, E. P. Dutton & Co. Inc. 1968.
- Fantino, E. y Duncan, B. Some effects of interreinforcement time upon choice, *Journal of Experimental Analysis of Behavior*, 1972, 17, páginas 3-14, 1972.
- Farina, A., Holzberg, J. D. y Dies, R. R. Influence of the parents and verbal reinforcement on the performance of schizophrenic patients, *Journal of Abnormal Psychology*, 1969, 74 (1), págs. 9-15.

- Fernández, G. (Comunicación personal), 1972.
- Fernández, *Revista Mexicana de la Investigación Psicológica*, 1969.
- Ferster, C. B. The use of the free operant in the analysis of behavior. *Psychological Bulletin*, 1953, 50, págs. 264-274. 1953.
- Ferster, C. B. y Skinner, B. F. *Schedules of reinforcement*, Appleton-Century-Crofts. 1957.
- Finney, D. J. *Experimental design and its statistical basis*, The University of Chicago Press, 1960.
- Fisher, R. A. *The design of experiments*, Hafner, 1966.
- Fixsen, D. L., Phillips, E. L. y Wolf, M. M. Achievement place: the reliability of self-reporting and peer-reporting and their effects on behavior, *JABA*, 1972, 5, págs. 19-30.
- Fjellestedt, N. y Sulzer-Azaroff, B. *Reducing the latency of a child's responding to instructions by means of a token system*, *JABA*, 1973, 6, páginas 125-130.
- Foreyt, J. P. y Kennedy, W. A. Treatment of overweight by aversion therapy, *Behavior Research and Therapy*, 1971, 9, págs. 29-34.
- Gaits, J. *Scale classification and statistics in statistical issues: a reader for the behavioral sciences*, Wadsworth Publishing Co. 1972.
- Gentile, J. R., Roden, A. H. y Klein, R. A. An analysis of variance model for the intrasubject replication design, *JABA*, 1972, 5, págs. 193-198.
- Giambra, L. M. Effect of number of irrelevant dimensions with ten concept types on the attribute identification task in the selection mode with exemplar and non-exemplar start cards, *Psychonomic Science*, 1969, vol. 14(2), 75-76.
- Glass, G. V. y Stanley, J. C. *Statistical methods in education and psychology*, Prentice-Hall, 1973.
- Glynn, E. L. Classroom applications of self-determined reinforcement. *Journal of Applied Behavior Analysis*, 1970, 3, págs. 123-132.
- Gorovitz, E. M. *The molecular basis of behavior*, Prentice-Hall, Inc. 1969.
- Gott, C. T. y Weiss, B. The development of fixed-ratio performance under the influence of ribonucleic acid, *Journal of the Experimental Analysis of Behavior*, 1972, 18, págs. 481-497.
- Grings, W. W., Schell, A. M. y Carey, C. A. Verbal control of an autonomic response in a due reversal situation, *Journal of Experimental Psychology*, 1973, vol. 99, núm. 2, págs. 215-221.
- Hake, D. F., y Campbell, R. L. Characteristics and response displacement effects of shock generated responding during negative reinforcement procedures: pre-shock responding and post-shock aggressive responding, *JABA*, 1972, 17, págs. 303-323.
- Hall, R. V., Lund, D. y Jackson, D. Effects of teacher attention on study behavior, *JABA*, 1968, vol. I. núm. 1, 1968, págs. 1-12.
- Hall R. V., Cristler, C., Cranston, S. S. y Tucker B. Teachers and parents as a researchers using multiple baseline designs. *JABA*, 1970, 3, págs. 247-255.
- Hall, R. V. *Managing behavior 1, behavior modification: the measurement of behavior*. H. & H. Enterprises, Inc. 1971.
- Hartmann, D. P. Forcing square pegs into round holes: some comments

- on "and analysis of variance model for the intrasubject replication design", *Journal of Applied Behavior Analysis*, 1974, 7, 4, 635-638.
- Hasazi, J. E. y Hasazi, S. E. Effects of teacher attention on digit-reversal behavior in an elementary school child, *JABA*, 1972, 5, págs. 157-162.
- Hays, W. L. *Statistics*, Holt, Rinehart & Winston, Inc. 1963.
- Hemmes, N. S. y Eckerman, D. A. Positive interaction (induction) in multiple variable interval, differential-reinforcement of high-rate schedules, *JEAB*, 1972, 17, págs. 51-57.
- Herbert E. W., Pinkston, E. M., Hayden, M. L., Sajnoway, T. E., Pinkston, S., Cordua, G. y Jackson, C. Adverse effects of differential parental attention, *JABA*, 1973, 6, págs. 15-30.
- Herman, L. M. y Arbut, W. R. Stimulus control and auditory discrimination learning seas in the bottle nose dolphin, *JEAB*, 1973, 19, págs. 379-394.
- Hernstein, R. J. Relative and absolute strength of response as a function of frequency of reinforcement. *JEAB*, 1961, 4, págs. 267-272.
- Hernstein, R. J. On the law of effect. *JEAB*, 1970, 13, págs. 242-266.
- Honig, W. K. *Operant behavior: areas of research and application*, Appleton-Century-Crofts, 1966.
- Holland, J. G. y Skinner, B. F. *Análisis de la conducta*, Editorial Trillas, 1970.
- Hope, K. *Manual práctico de estadística avanzada*. Editorial Trillas, México, 1970.
- Horner, D. R. Establishing use of crutches by a mentally retarded spina bifida child. *Journal of Applied Behavior Analysis*, 1971, 4, págs. 183-189.
- Hull, C. L. *A Behavior system*, Yale University Press. 1952.
- Hunt, E. B. Memory effects in concept learning. *Journal of Experimental Psychology*, 1961, vol. 62, núm. 6, págs. 598-604.
- Johnson, H. H. y Solso, R. L. *An introduction to experimental design in psychology: a case approach*, Harper & Row, Inc. 1971.
- Johnston, J. M. y Johnston, G. T. Modification of consonant speechsound articulation in young children, *JABA*, 1974, 5, págs. 233-246.
- Kazdin, A. E. y Bootzin, R. R. The token economy: an evaluative review, *JABA*, 1972, 5, págs. 343-372.
- Kempthorne, O. *The desing and analysis of experiments*, Wiley, 1952.
- Kendler, H. H., Glasman, L. D. y Ward, J. W. Verbal-labeling and cue-training in reversal-shift behavior, *Journal of Experimental Child Psychology*, 13, págs. 195-206. 1972.
- Keppel, G. *Design and analysis: a researcher's Handbook*, Prentice-Hall, 1970.
- Keselman, H. J., y Leventhal, L. Concerning the statistical procedures enumerated by Gentile, y otros. Another perspective, *Journal of applied Behavior Analysis*, 1974, 7, 4, 643-645.
- Killeen, P. A. A yoked-chamber comparison of concurrent and multiple schedules, *JEAB*, 1972, 18, págs. 13-22.
- Killeen, P. The matching law, *JEAB*, 1972, 17, págs. 489-495.
- Kirk, R. E. *Experimental design: procedures for the behavioral sciences*, Brooks/Cole Publishing Co. 1968.

- Kirk, R. E. (dir.). *Statistical issues: a reader for the behavioral sciences*, Wadsworth Publishing Co., 1972.
- Kish, L. *Some statistical problems in research design in statistical issues. A reader for the behavioral sciences*, Kirk, R. E. (dir.) 1972. Wadsworth Publishing Co. Inc.
- Koegel, R. L. y Covert, A. The relationship of self stimulation to learning in autistic children, *Journal of Applied Behavior Analysis*. 1972, 5, 381-387.
- Krontz, O. L. The separate worlds of operant and non-operant psychology, *Journal of Applied Behavior Analysis*, 1971, 4, 61-70.
- Krapfl, J. E., y Nawas, M. M. Client-therapist relationship factor in systematic desensitization, *Journal of Consulting and Clinical Psychology*, 1969, 33(4), págs. 435-439.
- Kratochwill, T., Alden, K., Demuth, D., Dawson, D., Panicucci, C. Arntson, P., McMurray, N. Hempstead, P. y Levin, J. A. Further consideration in the application of an analysis of variance model for the intrasubject replication design, *Journal of Applied Behavior Analysis*, 1974, 7, 4, 629-633.
- Krueger, L. E. Effect of irrelevant surrounding materials on speed of same-different judgment of two adjacent letters. *Journal of Experimental Psychology*, 1973, vol. 98, núm. 2, págs. 252-259.
- Lathrop, R. C. *Introduction to psychological research: logic, design, analysis*, Harper & Row, Publishers Inc. 1969.
- Loander, J. D. Effects of food deprivation on free-operant avoidance behavior, *JEAB*, 1973, 19, págs. 17-24.
- Liberman, R. P., Teingen, J., Paterson, R. y Baker, V. Reducing delusional speech in chronic, paranoid schizophrenics, *Journal of Applied Behavior Analysis*, vol. 6, núm. 1, 1973, págs. 57-64.
- Lindquist, E. F. *Design and analysis of experiments in psychology and education*, Houghton Mifflin, 1953.
- Livak, S. B. A comparison of two brief group behavior therapy techniques on the reduction of avoidance behavior, *Psychological Record*, 1969, 19, págs. 329-334.
- Logan, F. A. y Wagner, A. R. *Reward and punishment*, Allyn & Bacon, Inc. 1965.
- Lord, F. M. *On the statistical treatment of football numbers in statistical issues: a reader for the behavioral sciences*, Kirk, R. E. (dir.) Wadsworth Publishing, Co. 1972.
- Lovitt, T. C., Guppy, T. E. y Blattner, J. E. The use of a freetime contingency with fourth graders to increase spelling accuracy. *Behavior Research and Therapy*, 1969, 7, págs. 151-156.
- Luria, A. R. *A Mind of a mnemonist*. Basic Books, Inc. 1968.
- Mandelker, Annabel U., Bringham, T. A. y Bushell, D., Jr. The effects of token procedures on a teachers social contacts with her students. *JABA*, 1970, 3, págs. 169-174.
- Mandel, I. J. y Bridger, W. A. Cross-modality transfer of differential galvanic skin response conditioning to word stimuli, *Journal of Experimental Psychology*, 1973, vol. 99, núm. 2, págs. 157-161.

- Mann, R. A. The behavior-therapeutic use of contingency contracting to control an adult behavior problem: weight control, *JABA*, 1972, 5, págs. 99-110.
- Martínez N., (Artículo en preparación) 1975.
- Maxwell, A. E. *Experimental design in psychology and the medical sciences*. Wiley, 1958.
- McAllister, L. W., Stachowiak, J. G., Baer, D. M. y Conderman, L. The application of operant conditioning techniques in a secondary school classroom, *Journal of Applied Behavior Analysis*, 1969, 2, págs. 277-285.
- McGuigan, F. J. *Psicología experimental: enfoque metodológico*, Editorial Trillas, 1972.
- McKenzie, H. S., Clark, M., Wolf, M., Kothera, R. y Benson, C. Behavior modification of children with learning disabilities using grades as tokens and allowances as back-up reinforcers. *Exceptional Children*. 34, págs. 745-752, 1968.
- McNamara, R. J. Teacher and students as sources for behavior modification in the classroom, *Behavior Therapy*, 1971, 2, págs. 205-213.
- McNemar, Q. Sampling in psychological research. *Psychological Bulletin*, 1940, 37, págs. 331-365.
- McNemar, Q. *Psychological statistics*, John Wiley, 1969.
- Medland, M. B. y Stachnik, T. J. Good-behavior game: a replication and systematic analysis, *JABA*, 1972, 5, págs. 45-51.
- Meehl, P. E. Theory-testing in psychology and physics: a methodological paradox, *Philosophy of Science*. 1967.
- Meichenbaum, D. H. Bowers, K. S. y Ross, R. R. Modification of classroom behavior of institutionalized female adolescent offenders. *Behavior Research and Therapy*, 1968, 6, págs. 343-353.
- Meyer, J. B., Ströwing W. y Hosford, R. E. Behavior reinforcement counseling with rural high school youth, *Journal of Counseling Psychology*. 1970, 17 (2), págs. 127-132.
- Michael, J. Statistical inference for individual organism research mixed blessing or curse? *Journal of Applied Behavior Analysis*, 1974, 7, 4, págs. 647-653.
- Michael, J. Statistical inference for individual organism research: some reactions to a suggestion by Gentile, Roden, and Klein, *Journal of Applied Behavior Analysis*, 1974 a, 7, 4, 627-678.
- Milby, J. B. Modification of extreme social isolation by contingent social 149-152. *Journal of Applied Behavior Analysis*, 1970, 3.
- Millard, D. W. A conditioning treatment for "giggle micturition". *Behavior Research and Therapy*. 1966, 4, págs. 229-231.
- Moller, J. An experimental implementation of a role-governed response, tesis (inérita) de maestría. Universidad de Texas, 1973.
- Myers, J. L. *Fundamentals of experimental design*. Allyn & Bacon, Inc. 1972.
- Namoodiri, N. K. Experimental designs in which each subject is used repeatedly, *Psychological Bulletin*, 1972, vol. 77, núm. 3, págs. 54-64.
- National Geographic. Feb. 73, vol. 143, núm. 2.
- Nunnally, J. C. *Psychometric theory*, McGraw-Hill Book Co. 1967.

- O'Brien F., Bugle, C. y Azrin, N. H. Training and maintaining a retarded child's proper eating, *JABA*, 1972, 5, págs. 67-72.
- O'Leary, K. D. y Becker, W. C. Behavior modification of an adjustment class: a token reinforcement program, *Exceptional Children* 1967, 33, págs. 637-642.
- Panyan, M., Boozer, H. y Morris, M. Feedback to attendants as a reinforcer for applying operant techniques, *JABA*, 1970, 3, págs. 1-4.
- Perline, I. H. y Levinsky, D. Controlling maladaptive classroom behavior in the severely retarded, *American Journal of Mental Defficiency*, 1968, 73(1), págs. 74-78.
- Peterson, R. F. y Whitehurst, G. J. A variable influencing the performance of generalized imitative behaviors, *Journal of Applied Behavior Analysis*, vol. 4, núm. 1, 1971, págs. 1-9.
- Phillips, E. L., Phillips, E. A., Fixsen, D. L. y Wolf M. M. Behavior shaping works for delinquents, *Psychology Today*, junio, 1973. vol. 7, núm. 1, págs. 74-79.
- Pinkston, E. M., Reese, N. M., LeBlanc, J. M. y Baer, D. M. Independent control of a preschool child's aggression and peer interaction by contingent teacher attention, *JABA*, 1973, 8, págs. 115-124.
- Plutchik, R. *Foundations of experimental research*, Harper & Row, Publishers 1968.
- Pushking Publishing Group (dir.), *Psychology* 73/174, Enciclopedia, 1973.
- Quay, H., Werry, J., McQueen, M. y Sprague, R. Remediation of the conduct problem child in the special class setting, *Exceptional Children*, 1966, 32, págs. 309-315.
- Rachlin, H. C. On the tautology of the matching law, *JEAB*, 1971, 15, págs. 249-251.
- Rafi, A. A. Learning theory and the treatment of tics, *Journal at Psychosomatic Research*, 1962, 6, págs. 71-76.
- Resnick, J. H. Effects of stimulus satiation on the overlearned maladaptive response of cigarette smoking, *Journal of Consulting and Clinical Psychology*, 1968, 32(5), págs. 501-505.
- Restle, F. y Emmerids, D. Memory in concept attainment: effects of giving several problems concurrently, *Journal of Experimental Psychology*, 1966, vol. 71, núm. 6, págs. 794-799.
- Revushky, S. H. Some statistical treatments compatible with individual organism methodology, *Journal of the Experimental Analysis of Behavior*, 1967, 10, págs. 319-330.
- Ribes. E. *Técnicas de modificación de conducta: su aplicación al retardo en el desarrollo*, Editorial Trillas, México, 1972.
- Romanczyk, R. G., Kent, R. N., Diamant, C. y O'Leary, K. D. Measuring the reliability of observational data: a reactive process. *Journal of Applied Behavior Analysis*. Vol. 6, núm. 1, 1973, págs. 175-184.
- Rosner, S. R. Primacy in preschooler's short-term memory: the effects of repeated tests and shift-trials, *Journal of Experimental Child Psychology*, 13, págs. 220-230, 1972.
- Ryan, T. A. y Krumboltz, J. D. Effect of planned reinforcement counseling on client decision-making behavior, *Journal of Counseling Psychology*, 1964, 11, págs. 315-323.

- Santa, J. L. y Ranken, H. B. Effects of verbal coding on recognition memory, *Journal of Experimental Psychology*, 1972, vol. 93, núm. 2, págs. 268-278.
- Saunders, P. S. y Neimark, E. D. Effects of misinformative feedback and awareness of misinformation on concept attainment strategy. Proceedings, 76a. Convención Anual, APA, 1967.
- Schneider, B. A. y Meuringo, A. J. Responding under discretetrial fixed-interval schedules of reinforcement, *JEAB*, 1972, 18, págs. 187-199.
- Schwitzgebel, R. L. Preliminary socialization for psychotherapy of behavior disordered adolescents, *Journal of Consulting and Clinical Psychology*, 1969, 33, 1, págs. 71-77.
- Sheridan, C. L. *Fundamentals of experimental psychology*, Holt, Rinehart & Winston, Inc. 1971.
- Sidley, N. A. y Schoenfeld, W. N., Behavior stability and response rate as functions of reinforcement probability on "Random Ratio" Schedules, *Journal of the Experimental Analysis of Behavior*, 1964, 7, 281-283.
- Sidman, M. *Tactics of scientific research evaluating experimental data in psychology*, Basic Books, 1960.
- Sidowsky, J. B. y Smith, M. J. *Experimental methods and experimentation in psychology*, McGraw-Hill, 1966.
- Siegel, S. *Diseño experimental no paramétrico aplicado a las ciencias de la conducta*, Editorial Trillas, México, 1970.
- Sierra, G. P. y Castro, L. Efecto de las respuestas mediadoras de repaso y de la frecuencia del estímulo sobre la proporción de respuestas de aparcamiento retardado a las muestras, en las porciones de primacia y recencia, en las curvas de posición serial, (manuscrito inédito) 1975.
- Simón, H. A. *The sciences of the artificial*, The Massachusetts Institute of Technology, 1969.
- Skinner, B. F. *The behavior of organisms*, Appleton-Century-Crofts. 1938.
- Skinner, B. F. *Science and human behavior*, The Macmillan Co. 1953.
- Skinner, B. F. *Contingencies of reinforcement: a theoretical analysis*. Appleton-Century-Crofts. 1969.
- Skipper, J. K. Guenther, A. L. y Mass, G. *The sacredness of 05: a note concerning the uses of statistical levels of significance in social sciences*, Brooks/Cole. Kirk, R. E. (dir.), 1967.
- Stanley, J. C. An improving certain aspects of educational experimentation in improving experimental design and statistical analysis. Julian Stanley (dir.), Rand McNally. 1967.
- Steinberg, N. y Sekuler, R. Changes in visual spatial organization: response frequency equalization versus adaptation level. *Journal of Experimental Design*. 1973, vol. 98, núm. 2, págs. 264-251.
- Stevens, S. S. (dir.) *Handbook of Experimental Psychology*. John Wiley & Sons, Inc. 1951.
- Stice, J. E. (dir.) *The personalized systems of instruction (PSI)* College of Engineering. The University of Texas, Austin, 1971.
- Thoresen, C. E. y Elashoff, J. D. "An analysis of variance model for intrasubject replication design": some additional comments, *Journal of Applied Behavior Analysis*, 1974, 7, 4, 639-641.
- Tooley, J. T. y Pratt, S. An experimental procedure for the extinction of

- smoking behavior. *Psychological Record*. 1967, 17(2), págs. 209-218.
- Targerson, W. S. *Methods of sealing*. John Wiley & Sons, Inc. 1967.
- Ulrich, R., Stachnik, T. y Mabry, J. *Control de la conducta humana*. Editorial Trillas, 1972.
- Underwood, B. J. *Psicología experimental*. Editorial Trillas, 1972.
- Underwood, G. Control of selective attention and interference of processing in memory. *Journal of Experimental Psychology*. 1973. vol 99. número 1, págs. 28-34.
- Urbina, J., Gómez J., Vigil, S., Braverman, S. Sierra, G. y Castro, L. Control de los componentes de primacía y recencia mediante técnicas de repaso y de frecuencia de los estímulos, en una tarea de respuesta retardada en humanos, manuscrito inédito, 1975.
- Webb, E. J., Campbell, D. T. Schwartz, R. D. y Sechrest, L. *Unobstrusive measures: nonreactive research in the social sciences*. Rand McNally, 1966.
- Weis, L. y Hall, V. Modification of cigarette smoking through avoidance of punishment in Hall, V., *Managing Behavior (3), Behavior Modification: Applications in School and Home*, H. & H. Enterprises, 1971.
- Weisman, R. G. y Ramsden, M. Discrimination of a response independent component in a multiple schedule, *JEAB*, 1973, 19.
- Wincze, J. P., Leitenberg, H. y Agras, W. S. The effects of token reinforcement and feedback on the delusional verbal behavior of chronic paranoid schizophrenics, *JABA*, 1972, 5, págs. 247-262.
- Winer, B. J. *Statistical principles in experimental design*, McGraw-Hill Book Co. 1962.
- Whitlock, C. Note on reading acquisition: an extension of laboratory principles, *Journal of Experimental Child Psychology*, 1966, 3, págs. 83-85.
- Whittaker, J. O. (dir.) *Recent discoveries in psychology*, W. B. Saunders Company. 1972.
- Wolf, M., Risley, T. y Mees. Application of operant conditioning procedures to the behavior problems of an autistic child, *Behavior Research in Therapy*. 1964, 1, págs. 1-5-312.
- Yamane, T. *Statistics: an introductory analysis* (2a ed.), Harper & Row. 1970.
- Young, R. K. y Veldman, D. J. Heterogeneity and skewness in analysis of variance, *Perceptual and Motor Skills*, 1963, 16, págs. 588.

*Esta obra se terminó de Imprimir
el día 17 de Septiembre de 1990,
en los talleres de Editorial Nana, S. A.,
Av. Central núm. 62-A, Col. Tepalcates,
C.P. 09210, México, D. F.,
se encuadernó en Encuadernación Abeja,
Manuel Navarrete núm. 44, Col. Algarín,
C.P. 08688, México, D. F.,
se tiraron
1 000 ejemplares, más sobrantes de reposición*