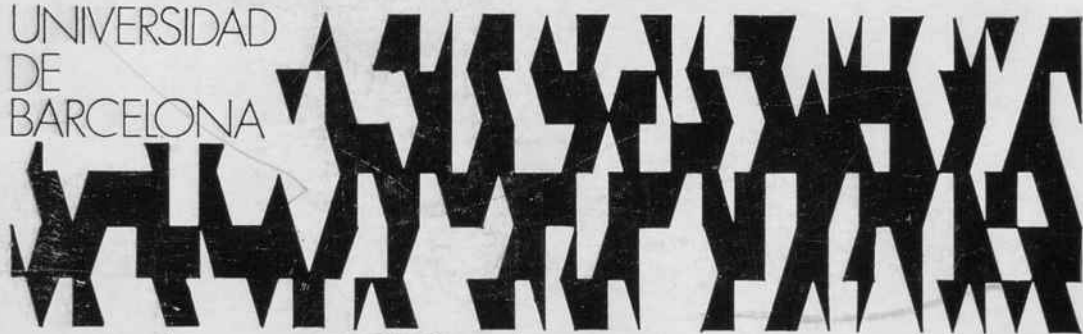


metodología científica aplicada a la investigación educativa

UNIVERSIDAD
DE
BARCELONA



instituto de ciencias de la educación

Informe n.º 10
Diciembre 1974

METODOLOGÍA CIENTÍFICA APLICADA A LA INVESTIGACIÓN EDUCATIVA

UNIVERSIDAD DE BARCELONA
INSTITUTO
DE CIENCIAS DE LA EDUCACION

Informe n.º 10
Diciembre 1974

METODOLOGIA CIENTIFICA
APLICADA A LA INVESTIGACION
EDUCATIVA

UNIVERSIDAD DE BARCELONA
INSTITUTO
DE CIENCIAS DE LA EDUCACION

Depósito legal: B. 2.056 - 1975

1975. — Ariel, S. A., Av. J. Antonio, 134-138, Esplugues de Llobregat - Barcelona

Sumario

Introducción	5
JAIME ARNAU	
La importancia de la observación en la investigación científica . . .	7
RAMÓN BAYÉS	
Presentación del método experimental	29
MARGARITA BARTOLOMÉ	
Estudio de las variables en la experimentación educativa	35
JAIME ARNAU	
El diseño experimental: su aplicación a las ciencias de la conducta .	61
CARLOS CUADRAS	
Modelos matemáticos del análisis de la varianza	91
PEDRO BATALLÉ	
Técnicas de muestreo	99
PILAR IBARROLA	
Pruebas de significación en el diseño experimental	117

INTRODUCCIÓN

El Instituto de Ciencias de la Educación de la Universidad de Barcelona organizó del 4 al 9 de marzo de 1974 un Seminario sobre: METODOLOGÍA CIENTÍFICA APLICADA A LA INVESTIGACIÓN EDUCATIVA, en el que participaron investigadores de los Institutos de Ciencias de la Educación y personas interesadas en el campo de la investigación educativa.

TEMARIO DEL SEMINARIO

Tema I. Métodos de investigación en función de la observación

- Estudio natural
- Estudio sistemático o diferencial
- Estudio experimental

Tema II. Metodología experimental

- Presentación del método
- Definición de las variables

Tema III. Presentación del diseño

- Diseño experimental
- Modelos matemáticos
- Pruebas de significación

Tema IV. Técnicas de muestreo

Tema V. Análisis de datos

Director del Seminario: JAIME ARNAU, Profesor de Psicología Experimental de la Universidad de Barcelona

Ponentes:

MARGARITA BARTOLOMÉ, Profesora de Pedagogía Experimental de la Universidad de Barcelona

PEDRO BATALLÉ, Jefe del Servicio de Estadística y Planificación del I.C.E. de la Universidad de Barcelona

RAMÓN BAYES, Profesor de Metodología Experimental de la Escuela Profesional de Psicología Clínica de la Universidad de Barcelona

CARLOS CUADRAS, Profesor de Biometría y Estadística de la Facultad de Ciencias de la Universidad de Barcelona

PILAR IBARROLA, Profesor Adjunto del Departamento de Estadística Matemática de la Facultad de Ciencias de la Universidad de Madrid

JOAQUÍN TORRENS IBERN, Catedrático de Estadística de la E.T.S.I.I. de la Universidad Politécnica de Barcelona

Barcelona, marzo de 1974

INTRODUCCION

El Instituto de Ciencias de la Educación de la Universidad de Buenos Aires organizó del 4 al 9 de marzo de 1971 un Seminario sobre Metodología Científica Aplicada a la Investigación Educativa, en el que participaron investigadores de los Institutos de Ciencias de la Educación y de otras unidades de ese campo de la investigación educativa.

Programa del Seminario

- Tema I. Metodología de investigación en función de la observación
Estadística actual
Estadística inferencial o hipotética
Estadística experimental
- Tema II. Metodología experimental
Formulación del método
Exposición de los variables
Forma III. Formulación del diseño
Diseño experimental
Método estadístico
Análisis de significación
- Tema IV. Formas de muestreo
Tema V. Estadística de datos
- Exposiciones del Seminario: Juan Añor, Exponer la Psicología Experimental de la Universidad de Buenos Aires
Mariano Kaminer, Evolución de la Psicología Experimental de la Universidad de Buenos Aires
Jorge del Sestizo de Estadística y Metodología del I.C.E. de la Universidad de Buenos Aires
Jorge Pérez, Evolución de la Metodología Experimental de la Escuela Psicológica de la Psicología Clínica de la Universidad de Buenos Aires
Carlos Correas, Evolución de la Historia y Estadística de la Facultad de Ciencias de la Universidad de Buenos Aires
Eduardo Añor, Evolución del Desarrollo de la Estadística Matemática de la Facultad de Ciencias de la Universidad de Buenos Aires
Jorge Añor, Estadística de la Facultad de la U.B.A. de la Facultad de Ciencias Exactas, Estadísticas de la Facultad de la U.B.A.
Exposición de la Facultad de Ciencias

Buenos Aires, marzo de 1971

LA IMPORTANCIA DE LA OBSERVACIÓN EN LA INVESTIGACIÓN CIENTÍFICA

por JAIME ARNAU

1. El papel de la observación en la labor investigadora
2. Condiciones de la observación científica
3. Problemas relativos a la observación
 3. a) ¿Qué deberá ser observado?
 3. b) ¿Cómo deberemos observar?
 3. c) ¿Qué técnicas de registro emplearemos?
4. Observación y control
 - 4.1. Métodos naturales
 - 4.2. Importancia de los métodos de observación natural
 - 4.2. a) La categorización
 - 4.2. b) Escalas de evaluación
 - 4.2. c) Sistemas de muestreo de la conducta
 - 4.3. Validez de las observaciones
 - 4.4. La fiabilidad de las observaciones
5. Métodos diferenciales o selectivos
 - 5.1. Métodos correlacionales
6. Métodos experimentales
 - 6.1. Experimento de campo
 - 6.2. Experimento de laboratorio
7. Experimento de laboratorio *versus* estudios de campo
8. Resumen
9. Bibliografía

I. EL PAPEL DE LA OBSERVACIÓN EN LA LABOR INVESTIGADORA

La experimentación, como cualquier otro tipo de investigación, está formada por un conjunto de actividades que tienden al descubrimiento y la verificación de fenómenos nuevos, con el objeto de conseguir esquemas teóricos significativos. En efecto, la *investigación científica* puede ser considerada como un "proceso", como una "secuencia de actividades" encaminadas a ampliar el ámbito de nuestros conocimientos. Ahora bien, podemos considerar este proceso, con el que caracterizamos la investigación científica, desde dos perspectivas diferentes. Por un lado, dicho proceso puede ser concebido como un conjunto de "fases o actividades" que constituyen cada uno de los pasos que determinan su estructura; y, por otro, puede ser interpretado como exponente de un determinado nivel de desarrollo dentro de un ámbito científico concreto. Desde este último punto de vista, el "proceso de investigación" representa los diversos estadios de desarrollo histórico de una ciencia; es decir, su "evolución". A su vez, considerado el "proceso de investigación" como un conjunto de actividades, influye decisivamente en la forma de cómo el científico va a planificar y orientar su investigación hacia el logro de unos conceptos significativos.

Suele ocurrir que estos dos aspectos, que la investigación presenta, se interrelacionen y combinen. De hecho, existen investigaciones que se mueven a niveles primarios en el desarrollo de una temática psicológica, y que, sin embargo, emplean métodos de observación altamente controlados y formalmente muy bien estructurados. Por el contrario, nos podemos encontrar con investigaciones que tengan niveles muy avanzados de desarrollo teórico y en las que se utilizan métodos de observación muy elementales y simples.

No obstante, hemos de reconocer que en todo estadio de desarrollo de una ciencia, ya sea en sus etapas iniciales o más avanzadas, se dan, en principio, las mismas fases o actividades que forman, de hecho, la estructura básica del proceso de investigación. Con la única diferencia que, en las etapas iniciales, estas actividades aparecen mucho menos sistematizadas y diferenciadas. A pesar de esto, como hemos señalado, en cualquier estadio de desarrollo en el que se presente la ciencia, las fases del proceso de investigación son, fundamentalmente, las siguientes: a) *formulación de un problema*, b) *recogida de datos*, c) *elaboración y análisis de los mismos*; d) *interpretación de los resultados*, y e) *la comunicación de los resultados*.

Ya se trate de los primeros estadios o de los más avanzados, el investi-

gador utiliza la "observación" como instrumento básico para el logro empírico de sus objetivos. De hecho, todas las ciencias empiezan con la observación, ya que, como señaló Cl. Bernard (1952), sólo a través de la observación se puede llegar a la obtención de conceptos con significación objetiva. Por esta razón la "observación" constituye uno de los aspectos más importantes del método científico.

En muchos casos, sin llegar a recurrir a métodos de observación muy elaborados como podrían ser "los experimentos de laboratorio", la ciencia ha conseguido una gran cantidad de información y conocimiento mediante el empleo sistemático de una cuidadosa observación. En muchas ocasiones los hombres de ciencia, llevados por su afán de un mejor conocimiento de las cosas, se han lanzado a la búsqueda, observación y descripción de fenómenos, y han aportado datos valiosísimos que han potenciado enormemente ulteriores investigaciones. Hemos de reconocer que las técnicas de observación se han perfeccionado enormemente durante los últimos años, lo cual ha permitido una investigación mucho más precisa y controlada.

Debido, pues, al especial papel que desempeña la observación en el proceso de la investigación científica, vamos a presentar las principales técnicas observacionales, fijando nuestra atención en la forma cómo cada una de ellas impone un determinado "nivel de restricción a nuestra observación". Asimismo podremos comprobar cómo, a medida que dichas técnicas se perfeccionan, cada una de las fases del "proceso de investigación" va adquiriendo una mayor diferenciación, derivándose los problemas no ya de simples intuiciones, sino de auténticas deducciones a partir de sistemas conceptuales coherentes. También constataremos cómo, a medida que la observación es mucho más controlada, se da una creciente utilización de instrumentos de medidas, de sistemas de control estandarizados, de análisis estadísticos cada vez más complicados, y una mayor sustitución de las simples clasificaciones cualitativas por mediciones cuantitativas mucho más exactas y precisas. Incluso en la forma de transmitir los resultados encontramos una enorme diferencia entre los procedimientos actuales y los anteriores. Así, actualmente, los reportes deben ceñirse a una serie de normas en cuanto a la presentación y disposición de los apartados, a una terminología básica, en definitiva, a un formato específico. Cada vez más, el método científico, en su incesante desarrollo y perfeccionamiento, se nos presenta como una técnica eficaz e imprescindible.

2. CONDICIONANTES DE LA OBSERVACIÓN CIENTÍFICA

Hemos destacado, en el punto anterior, la importancia que tiene la observación en todo proceso de investigación científica. Si bien hemos afirmado que toda investigación tiene su punto de arranque en la observación, no obstante, hemos de añadir que la consecución y formación de los conceptos

científicos depende, en gran parte, del rigor y la precisión con que dicha observación haya sido llevada a cabo. Ello exige al científico un constante perfeccionamiento de sus técnicas observacionales, así como la utilización de instrumentos y procedimientos de medida que le proporcionen unos datos cada vez más fiables y precisos.

Por otra parte, hemos de señalar que, desde un punto de vista general, son dos, como mínimo, las condiciones sobre las que descansa el proceso de conceptualización científico. Estas dos condiciones las encontramos virtualmente especificadas en la definición que nos da Campbell (1952) de la ciencia. Para Campbell la ciencia es "el estudio de aquellos juicios en relación a los cuales se puede lograr un acuerdo común". En esta definición, como hemos dicho, se encierran las dos condiciones básicas que informa el quehacer científico: *la objetividad y la verificabilidad*. Estas son las dos condiciones fundamentales que constituyen la garantía básica, el acuerdo universal entre observadores, que requiere el conocimiento científico.

Siguiendo, pues, la lógica de la definición que hemos dado de la ciencia, diremos que, es condición imprescindible en la presentación de cualquier fenómeno científico, la indicación, a modo de prescripción, de las "condiciones bajo las cuales se hayan realizado las observaciones". De esta forma, cualquier otro investigador podrá reproducir exactamente dichas condiciones y podrá observar, de nuevo, el mismo fenómeno. Con ello se consigue una garantía de su verificabilidad universal. Así, si un científico nos informa que: "sometiendo un determinado fenómeno X bajo las condiciones Y, se observará Z", tenemos la seguridad de que podremos de nuevo observar Z, al someter X bajo las condiciones Y. Por tanto, podemos concluir que el criterio de objetividad (y por consiguiente de verificabilidad) de un hecho depende del grado en que las condiciones bajo las cuales se produce hayan quedado "específicamente establecidas". Esto ofrece al científico la posibilidad de poder repetir las mismas observaciones siguiendo las prescripciones establecidas en el comunicado científico.

En efecto, todo proceso científico depende, en última instancia, de la objetivación de tales condiciones mediante el progresivo descubrimiento de los factores que pueden llegar a influir los hechos de nuestras observaciones.

Uno de los ideales de todo científico es alcanzar una "observación máximamente objetiva", para lo cual tiene que emplear una serie de procedimientos y técnicas que eliminen al máximo las posibles fuentes de distorsión y error. Por dicha razón, el progresivo desarrollo científico va unido a un control, cada vez mayor, de los factores distorsionantes. "En verdad, la historia de la psicología como ciencia ha sido el desarrollo de procedimientos e instrumentos de ayuda que, gradualmente, han eliminado y corregido las desviaciones y distorsiones en la realización de las observaciones" (Hayman, 1964).

En efecto, la elaboración y objetivación de unas técnicas e instrumentos

de medición constituye uno de los principales quehaceres del proceder científico. Ellas permitirán al investigador observar, registrar e interpretar, con las suficientes garantías, sus observaciones de modo que puedan verificarse en nuevas situaciones (H. Peak, 1953).

3. PROBLEMAS RELATIVOS A LA OBSERVACIÓN

Desde un punto de vista científico, cuando se pretenda llevar a cabo "observaciones" lo más objetivas posibles, deben plantearse, previamente, una serie de cuestiones de cuya solución dependerá la exactitud de las mismas. Entre las cuestiones más importantes están las relativas al tipo de procedimiento a utilizar, a la mayor o menor fiabilidad de las mismas, a la posibilidad de emplear nuevas técnicas, a la limitación de los fenómenos, etc. En síntesis, la problemática relativa a la observación puede sintetizarse en los siguientes puntos:

- a) ¿Qué deberá ser observado?
- b) ¿Cómo deberemos observar?
- c) ¿Qué técnicas de registro emplearemos?

Sólo después de que se hayan analizado y resuelto cada uno de estos tres aspectos, el científico podrá contar con una serie de "datos" a partir de los cuales podrá inferir conceptos con significado científico.

3. a) *¿Qué deberá ser observado?*

Toda observación tiene como finalidad la obtención de información sobre alguna cuestión concreta. Esto implica que antes de observar se debe tener alguna idea, aunque sea algo vaga, de lo que pretendemos observar. Sin duda alguna, esto nos obligará a una labor de selección. Hemos de señalar que, en esta labor de selección, nos será de mucha ayuda el poseer una previa "estructura teórica" o "esquema conceptual". En función, pues, de este marco conceptual, se orientará nuestra labor investigadora, y nuestro trabajo de selección de un material adecuado. Es decir, si tenemos el propósito de inferir, a partir de una serie de comportamientos, cuál es el tipo de proceso básico y la forma de cómo modifica la conducta, la selección del material a observar debe llevarse a cabo de acuerdo con unos presupuestos teóricos. Evidentemente, no es lo mismo estudiar el fenómeno de la agresión como mecanismo de defensa, tal como la concibe la teoría psicoanalítica, que como un impulso adquirido en una situación de frustración, como sugiere la teoría neo-conductista. En ambos casos, aunque se trata de un mismo hecho, las observaciones se orientarán de acuerdo al marco teórico, siendo éste el que, en definitiva, nos va a indicar qué conductas y qué situaciones deberemos observar a fin de confirmar las implicaciones establecidas por la teoría.

3. b) *¿Cómo deberemos observar?*

La observación, como hemos visto, no sólo obedece a un propósito teórico, sino que por otra parte se ajusta a unas determinadas condiciones. Toda observación debe ser realizada bajo unas "condiciones específicamente establecidas". Estas condiciones actúan a modo de reglas o prescripciones, las cuales establecen, de forma precisa, la naturaleza de la situación sobre la que se realizarán las observaciones y los procedimientos o instrumentos que deberán ser utilizados.

El análisis de estas condiciones, a las que ha de someterse nuestra observación, nos lleva a uno de los temas más íntimamente conexiónados con el problema que estamos discutiendo: *el control*. El control constituye una de las cuestiones clave en todo proceso de investigación, y particularmente, en la fase de obtención de datos. La palabra control dentro del ámbito de la investigación científica es sinónimo de "limitación o restricción". El control impone, pues, una serie de restricciones a nuestra observación. Restricciones que afectan tanto al científico observador (evitando que sus propios deseos o actitudes interfieran con los resultados), como a los propios sujetos cuya conducta es objeto de observación, así como a la propia situación que se presenta a los sujetos para que reaccionen.

En suma, referente a la cuestión de "¿cómo deberemos observar?" cabe destacar tres aspectos primordiales.

a) Necesidad de preparar un detallado análisis de las condiciones que constituirán la situación que va a ser presentada a los sujetos, a fin de observar cómo reaccionan a las mismas (instrucciones, tarea, ambiente, local, ruidos, etc.).

b) Prever el control de las posibles variables que sistemáticamente pueden alterar los resultados de nuestras observaciones.

c) Establecer, de un modo preciso, los procedimientos o técnicas que van a emplearse para que nuestras observaciones sean más fiables (aparatos, tests, cuestionarios, encuestas, etc.).

Todo esto supone, por tanto, una detallada descripción de la situación tal como ha de ser presentada al sujeto, así como una específica descripción de los principales determinantes que pueden afectar la conducta. Esto nos va a exigir la identificación concreta de las variables, así como el subsiguiente control aplicado a las mismas. Sólo así obtendremos datos que sean relevantes a nuestro problema. El olvido de estas implicaciones nos va a llevar, la mayoría de las veces, a interpretaciones erróneas de los fenómenos observados. Como ejemplo ilustrativo, tenemos el caso histórico del estudio de la inteligencia. Sólo pudo comprobarse la poca significación que tenían las medidas tomadas sobre dicho factor, cuando se demostró la poderosa influencia que sobre el rendimiento ejercían la motivación y los condicionamientos socio-económicos.

Hemos de señalar que cuando se habla de control frecuentemente se entiende a las limitaciones impuestas directamente sobre las situaciones (tipo de instrucciones, clase de estímulo, tarea, ejercicios, etc.). Ello tiene su justificación en el supuesto erróneo de que las variables que vamos a inferir son independientes y que, por tanto, su efecto sobre la conducta es constante y estable. Contrariamente, a dicha actitud extendida entre los investigadores, hemos de considerar que el control ha de extenderse a los propios sujetos experimentales, así como a los mismos investigadores, ya que como hemos indicado anteriormente, sus expectativas y deseos pueden llegar a modificar sus datos.

Es, pues, de suma importancia para la investigación clarificar el aspecto de “¿cómo va a observarse la conducta?”, a fin de conseguir una inferencia precisa acerca de lo que se está estudiando.

3. c) *¿Qué técnicas de registro emplearemos?*

Un último problema con respecto a la observación, nos lo plantea el “sistema de medida”.

Sin duda, la asignación de unidades métricas a nuestras observaciones posee un valor indiscutible ya que nos va a permitir manejar, de alguna forma, los datos brutos que hayamos obtenido de las mismas. Sin embargo, de una manera especial en el área de las ciencias del comportamiento, la medición plantea una serie de problemas graves. En primer lugar, se nos presenta el problema del “tipo de medida a utilizar”. ¿Clasificaremos nuestras observaciones por categorías, lo cual sólo implica el recuento de las mismas? ¿O es mejor, por otra parte, medir las diferencias e incluso la intensidad con que una característica conductual es poseída por los sujetos?

Supongamos que hemos decidido adoptar un sistema de medida adecuado a nuestros datos, consiguiendo un determinado nivel de cuantificación, en este caso nos podemos preguntar, también, por la significación de estas cantidades numéricas. Es decir, ¿las puedo sumar, multiplicar o correlacionar como me permite el modelo matemático? En otras palabras, si transformo unas observaciones a un sistema numérico determinado, con una serie de propiedades aritméticas, me puedo preguntar si el resultado de las operaciones que haya hecho con dichos números reflejará la naturaleza de los procesos psicológicos, o más bien será una consecuencia de las propiedades numéricas.

Cada uno de estos problemas deberán ser solucionados con la adopción de un adecuado sistema de medida que confiera a la cuantificación de las observaciones el rigor del sistema numérico y no modifique la significación de los fenómenos observados.

4. OBSERVACIÓN Y CONTROL

La condición fundamental para que nuestras observaciones se conviertan en "datos científicos" depende del grado de control que hayamos utilizado. Como principio básico podemos establecer que los datos científicos se originan de observaciones perfectamente controladas.

El control exige que nuestras observaciones se ajusten a unas implicaciones básicas. Dichas implicaciones imponen una serie de restricciones que afectan a los tres componentes integrantes de la situación que se investiga: a) el observador, b) el ambiente, y c) los sujetos observados.

Ahora bien, no siempre el control se suele aplicar de una forma estricta a cada uno de estos tres componentes. Así pues, según la amplitud que demos al control, tendremos tres métodos básicos de investigación, y que suelen dividirse en: 1) *métodos naturales*, 2) *métodos diferenciales o selectivos* y 3) *métodos experimentales*.

4.1. *Métodos naturales*

En los estudios naturales el observador no interviene activamente en la situación, ni modifica el curso natural de los procesos. Su papel es puramente pasivo, constatando, solamente, cómo se producen los fenómenos sin tratar de hacer nada que llegue a interferir los resultados de sus observaciones. Algunos autores (Campbell y Stanley, 1963) se han referido a dichos métodos como "estudios pre-experimentales".

Desde el punto de vista histórico la mayoría de los estudios biológicos realizados a lo largo del siglo pasado se ejecutaron de acuerdo con dicho método. Una cuidadosa observación de los animales y las plantas permitieron a los investigadores establecer unas características estructurales o morfológicas comunes. Este sistema proporcionó la base para la elaboración de clasificaciones o taxonomías.

Basándose en tales procedimientos, Darwin, en la segunda mitad del siglo pasado, pudo proponer su revolucionaria teoría de la evolución. Esta teoría ha servido para impulsar y estimular la investigación en campos tan amplios como la psicología comparada animal, la neurología, y la genética.

Concretamente en Europa, la corriente etológica ha ofrecido descripciones muy interesantes de conductas específicas animales, relativas a los hábitos de comida, conducta de apareamiento, maternidad y agresión. Estos estudios implicaron observaciones sistemáticas de las reacciones animales en situaciones naturales. Gracias a esta detallada información se cuenta con una base interpretativa para los resultados de los experimentos de laboratorio. Muchos de los estudios naturales realizados con ratas y monos han aportado nuevos conocimientos sobre su estructura social, y han permitido interpretar

adecuadamente algunos experimentos de laboratorio. Una variante del método natural lo constituye el *método clínico*. Éste se fundamenta, básicamente, en las observaciones realizadas por un terapeuta en el curso de una entrevista y tratamiento del paciente, en relación a problemas psicológicos. El médico alemán Emil Kraepelin, usando tan sólo la descripción y la clasificación, propuso dos grandes categorías de las enfermedades mentales: la psicosis maníaco-depresiva y la esquizofrenia. Sistema todavía vigente en el momento de dictaminar legalmente la enfermedad mental. Todos los estudios experimentales psiquiátricos toman, como punto de arranque, dicha clasificación. También Freud utilizó el método clínico. Él nos presentó un detallado análisis y descripción de la diferencia entre las parálisis histéricas y las de origen orgánico. Freud pudo deducir que los síntomas histéricos tenían su origen en fuertes bloqueos emocionales asociados con los órganos paralizados del cuerpo, y que, por consiguiente, toda terapia tendría que orientarse en romper tal asociación.

Otro de los ejemplos importantes, dentro del campo de la psicología en la aplicación del método natural, lo tenemos en el trabajo del psicólogo suizo Jean Piaget. Jean Piaget ha dedicado toda su obra al estudio del desarrollo del pensamiento y el lenguaje en el niño, empleando para ello un método de observación que él mismo lo ha definido como "método clínico". Los estudios de Piaget están precedidos por observaciones naturales del comportamiento infantil en diversas situaciones. A partir del análisis y descripción del comportamiento infantil, pudo llegar a establecer diversas categorías o estructuras mentales, así como una serie de situaciones prueba y preguntas estándar que podrá aplicar a los niños en ulteriores investigaciones.

El carácter propio de las observaciones naturales y clínicas es reducir al mínimo las restricciones sobre la observación y la situación observada. De esta manera, la intervención deliberada, mediante manipulación directa, se evita al máximo debido a las interrupciones que supondrían en el desarrollo espontáneo del comportamiento, alterando totalmente el proceso conductual.

4.2. *Importancia de los métodos de observación natural*

La importancia de los estudios de observación natural, llamados, también, estudios de campo, debido a que la conducta ocurre fuera del contexto del laboratorio, estriba en el hecho que son una fuente inagotable de nuevas ideas. La ciencia utiliza primordialmente la observación cuando se lanza a la búsqueda de nuevos conceptos, más bien que cuando necesita una confirmación más precisa de los mismos. También de la observación se originan una gran cantidad de problemas, que van a constituir la base para futuras investigaciones experimentales. En suma, los métodos de observación natural constituyen el instrumento de trabajo propio de los estadios iniciales de las ciencias, en donde la creatividad, la imaginación y capacidad organizativa del investigador juegan un papel muy importante.

No obstante, los métodos observacionales juegan un papel importante en el concierto de la metodología científica, y poseen sus propias técnicas para la descripción y recogida de datos. Estas técnicas, como la de los métodos más perfeccionados, tienen por finalidad el descubrimiento de las relaciones existentes entre los diversos fenómenos naturales. Y aunque el investigador tiene como norma básica, en este tipo de estudios, la de no interferir con los hechos, puede servirse de una serie de procedimientos no sólo para observar adecuadamente los fenómenos, sino también para registrarlos de una forma precisa y conservarlos para su ulterior estudio.

Debido, pues, al amplio uso que han tenido los métodos de observación, sobre todo en determinadas áreas del comportamiento, algunos de ellos han adquirido un gran desarrollo a lo largo de estos últimos años. Podemos señalar como sistemas de observación más importantes los siguientes: a) *la categorización*, b) *la escala de evaluación*, y c) *el muestreo de conducta*.

4.2. a) La categorización

La categorización es una de las formas más útiles para recoger las observaciones realizadas sobre la conducta. La categoría viene determinada por una expresión concreta de la conducta y abarca a un conjunto de conductas de características similares. De esta forma, los sistemas de categorías proporcionan al investigador un adecuado marco de referencia que le permite no sólo recoger sino también organizar sistemáticamente los datos. La categorización es, pues, un procedimiento para manejar el flujo continuo de acontecimientos conductuales según un criterio sistemático y ordenado.

Al tratar, pues, de los sistemas de categorías hemos de tener en cuenta que, éstas han de establecerse previamente por el investigador, de acuerdo con unos presupuestos básicos. Estos presupuestos dependen del nivel de conceptualización conseguido, de la cantidad de los fenómenos observados, de los objetivos de la investigación, etc. El sistema de categorización depende, en suma, de las exigencias teóricas y prácticas de la investigación.

Uno de los criterios más frecuentemente utilizados para la formación de categorías es el de la "exhaustividad". Según este criterio cualquier segmento de conducta observable deberá poder ser clasificado en alguna de las categorías que un sujeto podrá adoptar ante una situación de resolución de problemas; es exhaustiva desde el momento que se ha previsto de antemano las posibles formaciones de reacciones de los individuos.

Tenemos, por otra parte, los sistemas de categorías "no-exhaustivos"; aunque un análisis más profundo de los mismos nos llevaría a deducir que no son sino una variedad de los anteriores. En otras palabras, un sistema de categorización "no-exhaustivo" es "exhaustivo" desde el momento que registra todas aquellas manifestaciones de conducta que no concuerdan con ninguna de las categorías establecidas de antemano, en una categoría común denominada "fuera del sistema".

A pesar de que, aparentemente, ambos sistemas parecen ser idénticos, desde un punto de vista práctico existen toda una serie de consideraciones de tipo económico que nos permiten afirmar que el sistema "no-exhaustivo" es considerablemente más rápido, con el consecuente gran ahorro de esfuerzo y tiempo.

Un segundo criterio que podemos adoptar al formar un sistema de categorías se refiere al *grado de inferencia* que deberá realizar el científico u observador. Así, algunos de los sistemas sólo requieren que los observadores registren los segmentos conductuales, tal como éstos se manifiestan externamente (nivel fenotípico). Es decir, sin exigírseles ningún tipo de inferencia. Opuestos a éstos encontramos los sistemas que requieren, por parte del observador, la codificación de la conducta en función de las deducciones realizadas a partir de su manifestación extensa (nivel genotípico). Así por ejemplo, supongamos que un investigador se halla interesado por el análisis del comportamiento infantil en el juego, y se dedica a codificar observaciones tales como: "Juan empuja a sus compañeros", "Pedro se ha negado a colaborar" y "José se deja pegar", en este caso el investigador se movería a nivel fenotípico. Ahora bien, si frente a las mismas observaciones, clasificara las conductas de acuerdo con las siguientes expresiones: "Juan se ha mostrado hostil", "Pedro no es sociable" y "José es sumiso", ha categorizado a partir de las inferencias obtenidas de sus observaciones. En este segundo caso, se mueve a nivel genotípico. En la primera situación, el observador registra la frecuencia con que se presentaban unos hechos, dejando para más tarde, en el análisis de los datos, la interpretación de los mismos. En la segunda situación, el investigador realiza las inferencias en el mismo momento que observa.

El principal problema que deberemos resolver, en el caso que decidamos adoptar el sistema de categorización inferencial, es que se establezca de antemano, y de forma precisa, qué tipo de conducta refleja un determinado estado subjetivo del individuo. Es decir, qué clase de conducta fenotípica se mueve en una misma dimensión genotípica. En el caso de que no existiera un claro acuerdo previo, es mejor utilizar el sistema de registro de conducta directa, dejando para más tarde la interpretación de la misma.

Por último se pueden formar categorías siguiendo el criterio de la "continuidad" o "no-continuidad" del sistema. Si el criterio que adoptamos nos permite categorizar las conductas a lo largo de un continuo de valores, tendremos un tipo de categorización "continua". Esto implica que los fenómenos deberán mantener entre sí cierta relación. Si por el contrario no existe esta relación dimensional entre los hechos observados, podemos categorizarlos según categorías discretas. Si deseamos categorizar la agresividad, podríamos adoptar un sistema de tres categorías siguiendo un criterio de menor a mayor intensidad en la manifestación de la misma. Así, podríamos establecer las siguientes categorías: primera categoría "insulto verbal", segunda categoría "amenaza verbal", y tercera, "ataque físico". Según este criterio las categorías

se ordenan en una escala continua, en la que las diversas clases representan distintos niveles de manifestación de la agresividad.

Cuando entre las diferentes clases no existe ningún tipo de relación, tendremos que adoptar el sistema de categorización descrito.

4.2. b) Escalas de evaluación

Podemos utilizar, también, la escala de evaluación como sistema de recogida de datos. En líneas generales podemos señalar que este procedimiento es, en principio, menos fiable y menos consistente que un sistema bien estructurado de categorías. Ahora bien, frecuentemente, por exigencias prácticas, nos vemos obligados a utilizar este sistema de recogida para la obtención de datos.

Se entiende por *escala de evaluación* a un conjunto de valores que describen diversos grados de posesión de una característica psicológica por parte de los sujetos. Es decir, representa la observación de una de las dimensiones de la conducta. La utilización de dichas escalas pueden tener dos formas: 1) Podemos registrar la conducta a intervalos diferentes. Así, podemos utilizar la escala antes y después de que haya ocurrido un determinado fenómeno. Según esto podemos evaluar, a título de ejemplo, el grado de comunicación de un grupo antes y después de haber sido sometido a una "dinámica de grupos". Evidentemente constataríamos unas notables diferencias entre ambas escalas. 2) Podemos, también, evaluar un hecho después de que éste se haya producido totalmente. En el ejemplo anterior, podríamos tomar una serie de medidas sobre aspectos tales como la agresividad, ansiedad, dependencia, etc., después de que el grupo haya finalizado todas las sesiones.

Una de las principales ventajas que nos ofrece el sistema de evaluación es que nos proporciona los datos, de alguna manera, ya cuantificados. Esto supone un notable ahorro de tiempo y trabajo. Es obvio que para el mismo caso se pueden utilizar sistemas de categorización, no obstante, hemos de tener en cuenta que si los indicios pertenecientes a cada una de las categorías abarca una amplia variedad de conducta, su registro nos llevaría a una intensa labor de observación. Esto se puede simplificar empleando el método de escalas de evaluación, el cual nos ofrece un sistema mucho más eficaz y simplificado, tanto en la obtención de datos, como en la evaluación de diversas dimensiones conductuales.

Si bien, como hemos dicho, el sistema de evaluación constituye un método práctico de observación, desde el punto de la fiabilidad parece ser menos preciso. Ello depende, en gran parte, que los juicios de valor que deberá realizar el observador sobre los segmentos de conducta no tienen la suficiente garantía para asegurarnos que la valoración haya sido exacta. Por esta razón se aconseja, siempre que sea practicable, la utilización del sistema de cate-

gorias cuando estamos interesados en lograr una mayor precisión en nuestras observaciones.

4.2. c) Sistemas de muestreo de la conducta

La metodología científica, ha desarrollado diversos sistemas de observación para obtener datos de aspectos tan complejos como la conducta humana. En muchos casos la conducta de los individuos se presenta como un conjunto de interacciones en lapsus de tiempo tan rápido que difícilmente se pueden observar y codificar. A tal efecto, Wright (1960) nos presenta unos sistemas de muestreo de conducta, para el estudio del comportamiento infantil, que nos puede servir de pauta para el empleo de este sistema de observación.

Los cuatro sistemas de muestreo son los siguientes:

a) *Las descripciones diarias*, en las que, mediante el empleo de instrumentos tales como cámaras cinematográficas o cintas magnetofónicas, el investigador registra largas secuencias conductuales. De esta manera se logra acumular una gran cantidad de datos que van a constuir su material de trabajo.

b) *Descripciones de conducta concretas*. Estas descripciones se ciñen a fragmentos muy concretos de la conducta que se presentan en situaciones tipificadas. Por ejemplo, cuando se desea observar el comportamiento del niño ante la presencia de un juguete raro.

c) *Muestreo del tiempo*. Éste es, sin duda, uno de los sistemas más eficaces para la obtención de muestras representativas de la conducta observada. Mediante esta técnica se selecciona un determinado intervalo de tiempo el cual se aprovecha para efectuar las observaciones. Los intervalos pueden variar desde varios segundos a veinte minutos.

d) Y por último tenemos el *muestreo segmental*, en el que se selecciona una actividad determinada, como por ejemplo reacciones de temor, actos agresivos, etc., registrándose las veces que aparecen en un determinado intervalo de tiempo.

Wright añade que los diversos sistemas de muestreo de la conducta no difieren en relación a su confiabilidad, llegándose a comprobar que todos ellos presentan un mismo índice de acuerdo entre observadores. Hemos de insistir, de nuevo, en la validez que poseen estas técnicas para el estudio de la conducta en sus más diversas manifestaciones.

4.3. Validez de las observaciones

El problema de la validez de las observaciones ha constituido un aspecto que, frecuentemente, ha sido olvidado, en el momento de valorar los sistemas de observación.

Para Stouffer (1950) el problema de la validez de nuestras observaciones es, en definitiva, un problema de predicción. En la medida que nuestros

datos concuerdan con criterios externos a los que se hallan directamente vinculados, contaremos con una base para argumentar sobre su validez. Podemos también llegar a comprobar la validez de las observaciones por vía de la deducción teórica. Como veremos, a medida que las diversas implicaciones derivables de una estructura teórica reciben una confirmación práctica en nuestros datos, éstos son cada vez más válidos. De acuerdo con lo expuesto, contamos, en principio, con dos principales sistemas de validación:

a) El primero consiste en comprobar hasta qué punto existe un acuerdo o relación entre las observaciones realizadas sobre una variable y las medidas obtenidas de la misma, cuando se ha empleado un sistema de medida independiente. Este sistema recibe el nombre de *criterio*. El empleo de este procedimiento se considera como una forma de validación externa.

En este caso, el problema más importante es encontrar el criterio. "En realidad, la selección de cualquier criterio se basa en una teoría acerca de la relación existente entre el proceso medio y la situación o conducta-criterio, pero por lo general la teoría implícita es algún supuesto común, tal como la noción de que una persona a quien le disgusten los negros tenderá a votar por la segregación más que otra que los vea con agrado, o la presunción de que es más probable que vaya a la iglesia quien tiene actitud favorable hacia ella que quien exprese un amargo resentimiento en su contra" (H. Peak, 1953).

A medida que se confirma la correlación existente entre el criterio y los datos observados, éstos van adquiriendo mayor validez y significación objetiva. Sin embargo, el principal caballo de batalla, como ya hemos indicado, es encontrar un sistema independiente de medida que "por hipótesis" esté altamente relacionado con nuestra variable.

b) Un segundo procedimiento de validación depende del grado en que nuestras observaciones cumplen con las implicaciones derivadas de la teoría. En este caso, validar las medidas u observaciones consistiría en verificar en ellas las predicciones establecidas por la teoría.

Si la teoría establece que la conducta aprendida depende de un ejercicio repetitivo y tales predicciones resultan correctas, tanto la teoría como nuestra medida resultan beneficiadas. En caso contrario, nos deberíamos plantear el problema de revisar la teoría o modificar nuestros instrumentos de observación. Si probada la teoría en nuevas situaciones se cumplen sus predicciones, estaremos obligados a modificar nuestros sistemas de observación y medida.

El criterio de validación basado en la verificación de la predicción teórica no deja de ser un tanto problemático, debido a la gran cantidad de fuentes de error, que pueden no sólo falsear la situación de modo que las implicaciones teóricas no se cumplan, sino incluso, distorsionar nuestras observaciones. Ésta es una de las principales razones por la que se tienda a emplear sistemas externos de validación.

Pensar que la confirmación de una predicción, implica una verificación automática de la teoría y de nuestros instrumentos de medidas, es una pre-

tensión un tanto confiada e ingenua. En todo caso, siempre que obtengamos una prueba sobre la validez de unas observaciones, concluiremos que sólo es atribuible a una situación concreta, y que cualquier otra aplicación de las mismas deberá ir acompañada por la presencia de unas condiciones que aseguren la igualdad de la situación.

4.4. *La fiabilidad de las observaciones*

Si la validez es un problema de significación de observaciones, la fiabilidad descansa de un principio o supuesto implícito según el cual nuestras observaciones serán estables siempre que se den las mismas condiciones de producción de los fenómenos. Este criterio es fundamental para la buena marcha de la labor científica, y establece, como norma básica, *que las condiciones que hacen variar la conducta sean siempre las mismas si queremos obtener las observaciones similares*. Si éstas se alteran, consecuentemente será debido a un cambio en las condiciones determinantes de la conducta.

Según Loevinger (1947) son dos las causas responsables de la inestabilidad de nuestras observaciones. Una es debida a la incapacidad, por parte de los estímulos, de suscitar las mismas reacciones en situaciones diferentes, y la otra se produce cuando entre la situación-estímulo y la respuesta se interfieren una serie de procesos, como las actitudes, los motivos, etc., que pueden llegar a alterar los fenómenos observados. A estas dos fuentes de error, H. Peak (1953) añade dos más: *a)* la presencia de nuevos estímulos que pueden llegar a afectar las respuestas, y *b)* las variaciones que pueden producirse en el registro e interpretación de los fenómenos observados. La única forma que disponemos para eliminar este último tipo de error, es calculando, de alguna forma, el grado de acuerdo entre las medidas de varios observadores. Para la reducción o eliminación de las restantes fuentes de error podemos aplicar las técnicas de control experimental.

En conclusión, podemos afirmar que la fiabilidad no es una propiedad que depende exclusivamente del mayor o menor acuerdo que se consigue entre una serie de observaciones sobre un mismo fenómeno, sino también, en gran parte, de la constancia de todos los determinantes de que una manera u otra pueden influir en la conducta. De ahí, que el grado de estabilidad que pueda conseguirse entre las diferentes observaciones, cuando éstas se han repetido, depende del grado de control que se haya conseguido de los factores que pueden llegar a alterar nuestras observaciones.

5. MÉTODOS DIFERENCIALES O SELECTIVOS

Estos procedimientos aprovechan las diferencias existentes entre los propios datos como sistema de clasificación de las observaciones. Si nos interesa, por ejemplo, estudiar la influencia del sexo sobre los niveles de aspiración, mediremos estos niveles a partir de muestras diferenciadas de una forma

natural. Es decir, la variable independiente de nuestro estudio, el sexo, se presenta dicotomizada en la naturaleza, antes que el investigador intervenga.

Uno de los rasgos diferenciadores más importantes entre estos métodos y los experimentales, es que, en estos últimos, las condiciones de producción de los fenómenos son manipuladas directamente por el investigador, mientras que en los primeros, el investigador aprovecha las condiciones tal como se encuentran.

Algunas de estas condiciones suelen presentarse en forma de categorías cualitativas, como el sexo, la edad, clase social, diagnóstico psiquiátrico, etc. Por otra parte, también, se pueden conseguir la diferenciación de los sujetos mediante los puntajes obtenidos en un determinado tipo de test. La característica medida por el test queda expresada en una dimensión básica cuyos valores extremos nos sirven para formar dicotomías. Ejemplo, sujetos neuróticos y no neuróticos, introvertidos y extrvertidos, ansiosos y no ansiosos, etc. Formados los grupos en base a estas dicotomías, podemos esperar que cualquier cambio que se opere en la conducta observada de los individuos sea efecto de esta categorización previa.

La actual investigación piagetiana se orienta por la combinación de los métodos clínicos con los diferenciales. Así, de esta manera, Piaget utiliza, inicialmente, el método clínico para obtener algún conocimiento sobre la estructura del proceso del pensamiento del niño, y luego intenta comprobar las variaciones en la estructura del pensamiento observando al niño en función de la edad. La razón básica por la que el método empleado por Piaget se denomine clínico y no diferencial, estriba en que sus procedimientos no nos permiten reproducir y verificar sus observaciones de una forma estandarizada. Sus observaciones no son susceptibles de una adecuada cuantificación, que nos sirva para establecer de un modo adecuado la variación del rendimiento para las diferentes edades del niño. Hemos de tener en cuenta que el método diferencial permite la cuantificación de las observaciones y consiguientemente, su objetivación. Un intento de aplicación del método diferencial a la temática piagetiana lo encontramos en el trabajo de Goodnow. Goodnow (1962) utilizó, en su estudio, tareas muy similares a las de Piaget, como los juicios de peso, volumen y espacio y las estandarizó. Es decir, elaboró un conjunto de instrucciones, construyó una serie de piezas de idéntica forma, etc., Goodnow nos presenta en su trabajo una especificación concreta de tareas e instrucciones, de tal forma que quien pretenda repetir su experimento pueda llevarlo a cabo, con la seguridad de que las condiciones serán exactamente las mismas.

5.1. *Métodos correlacionales*

Como una subclase del método diferencial, podemos emplear la técnica correlacional. Dicha técnica tiene como objetivo principal el de comprobar la relación existente entre dos o más hechos observados o medidos.

Mediante el método correlacional se pretende descubrir las relaciones que tienen los hechos psicológicos, tal como se presentan en la naturaleza. Es decir, nos permite el conocimiento de las posibles conexiones entre los fenómenos psicológicos. Para ello el investigador puede utilizar la variable, sin que haga falta una previa manipulación de la misma.

Otra de las características que presentan los estudios correlacionales, es la falta de criterio fijo para la elección de la variable independiente y dependiente. Más bien el hecho de que una variable, en un estudio dado, se considere como independiente o dependiente, es una cuestión de tipo arbitrario y depende, fundamentalmente, de los propósitos que persigue la investigación. Si podemos constatar que la relación entre las dos variables es muy alta, contamos con una buena base para poder preceder las futuras variaciones de las variables, a medida que aumenta nuestro conocimiento de los elementos covariantes.

A pesar de las dificultades que supone el empleo del método correlacional no sólo en relación al establecimiento de relaciones entre variables, sino también con respecto a la selección de sujetos, al control de los factores distorsionantes, etc., estos métodos constituyen una valiosa ayuda y nos aportan una información estimable en áreas donde la investigación experimental es difícil de aplicar. Sobre todo se ha demostrado la efectividad de tales procedimientos en estudios relacionados con el aprendizaje, cambios de opinión, efectos de "mass media", cambios de actitud, etc.

Una última consideración de los métodos diferenciales, en general, nos permite afirmar que ellos no sólo constituyen el puente de unión entre los procedimientos de observación natural y los estudios experimentales, sino que constituyen una inagotable fuente de sugerencias e hipótesis susceptibles de verificación experimental.

6. MÉTODOS EXPERIMENTALES

Se conoce por experimento aquella situación en la que el investigador introduce un estímulo o modifica alguno de los componentes de aquélla, y, a continuación, observa cómo reacciona el individuo. Esto nos permite llegar a la conclusión de que lo que se pretende con la técnica experimental es provocar la respuesta del sujeto. Cuando sometemos un individuo a un prueba o le pedimos que realice una tarea, estamos buscando su posible respuesta a fin de convertirla en dato.

La forma como el experimentador puede llegar al total dominio de la situación, es mediante el control. En la experimentación el investigador posee un control directo de la variable independiente, la manipula, determina sus valores, en suma, depende totalmente de él.

Para entender mejor cómo funciona un experimento, vamos a proponer un ejemplo. Supongamos que queremos estudiar la influencia de la "motivación"

sobre la "capacidad discriminativa" de los animales. Para ello formamos tres grupos de ratas y las sometemos, respectivamente a una, ocho y dieciséis horas de privación de alimento. Podríamos, incluso, formar un cuarto grupo, para conseguir una línea de base o criterio de comparación. En el caso de que el experimentador crea que existen variables independientes que deban ser controladas, podrá utilizar cualquier técnica de control experimental, con objeto de tener la seguridad que las únicas condiciones de variación son las que han sido impuestas por él. Sólo así posee un argumento y una base para la verificación de su hipótesis.

Hemos de señalar que el experimentador, antes de poner en marcha su proyecto de investigación, deberá haber resuelto una serie de puntos importantes. Así, concretamente, tiene que haber tomado una decisión sobre los valores que va a utilizar de su variable independiente, de cómo va a formar los grupos, sobre el tamaño de los mismos, etc. Todas estas decisiones, que deben ser tomadas antes de iniciar el trabajo, constituye lo que se conoce con el nombre de *planificación experimental*. También suele recibir el nombre de *diseño experimental*. Cada experimento supone, pues, un diseño propio, que nos va a permitir resolver una serie de cuestiones con el objeto de obtener un máximo de información con un mínimo de coste. Como veremos más adelante, existen en la práctica una serie de diseños totalmente planificados, e incluso etiquetados, que en muchos casos pueden facilitar enormemente nuestra labor planificadora. Ahora bien, antes de decidimos por la utilización de uno de estos diseños hemos de tener en cuenta una serie de cuestiones, tales como la cantidad de información que puede obtenerse por unidad de observación, la cantidad de variables que vale la pena estudiar simultáneamente, si el azar es la mejor estrategia o no, etc.

En suma, podemos afirmar que el principio básico de la experimentación consiste en actuar directamente sobre las condiciones que influyen directamente sobre los sujetos, con el fin de observar de qué manera su conducta acusa la acción de tales condiciones. Esto implica un conocimiento de cómo se habrían comportado en el caso de que no hubiesen existido tales condiciones. Este constituye, en definitiva, nuestro criterio de comparación, para poder así inferir la eficacia de nuestros tratamientos.

Expuestos los principios básicos de la experimentación, vamos a estudiar a continuación sus formas más importantes: *el experimento de campo y el experimento de laboratorio*.

6.1. *Experimentos de campo*

Hemos afirmado anteriormente que la experimentación puede diversificarse en dos grandes métodos, los experimentos de campo y los de laboratorio (experimento en el sentido más estricto de la palabra). Con el experimento de campo nos encontramos con un tipo de investigación que constituye el

lazo de unión entre los llamados estudios de campo (así podrían denominarse, en principio, los métodos estudiados hasta ahora), y el experimento propiamente dicho.

El experimento de campo, a diferencia de los estudios de campo, implica la manipulación directa de las condiciones experimentales. En los estudios de campo el investigador puede seleccionar los sujetos y medir las condiciones existentes en el campo, a fin de establecer o descubrir posibles relaciones. En el experimento de campo, el investigador actúa directamente sobre la variable independiente, aunque en muchos casos sea difícil de lograrlo, y no espera a que se produzca naturalmente. El experimento de campo exige, pues, que las condiciones se hayan preparado de antemano.

¿Qué ventajas nos reporta el experimento de campo? Con el experimento de campo se estudian los problemas en contextos reales. Se trata, pues, de evitar las situaciones artificiales que se crean en los laboratorios. Con ello se pretende que los individuos reaccionen en condiciones lo más posiblemente naturales y que su conducta refleje una actuación en una situación real. Se ha podido comprobar que los sujetos, al ser estudiados dentro de un laboratorio, aceptan el papel de "sujeto de estudio", lo cual puede modificar totalmente la situación. Y no sólo existe esta grave dificultad, sino que además, la situación de laboratorio supone la aplicación de una serie de restricciones con lo que la posibilidad de generalizar los resultados queda totalmente limitada.

En virtud, pues, de lo dicho no podemos cuestionar, a continuación, la artificialidad, incluso, de los experimentos de campo. Desde un punto de vista estricto, si consideramos que en un experimento de campo estamos aplicando unos controles, sometemos a los sujetos a unos determinados cuidados, les damos instrucciones, etc., ¿no convertimos, también, la situación de campo en una situación artificial? Esta consideración nos impide establecer una línea divisoria tajante entre lo "artificial" como propio del trabajo de laboratorio, y lo "real" como propio de los trabajos de campo. Parece más bien que entre estos dos conceptos existen diferencias de grado.

6.2. *Experimento de laboratorio*

El experimento de laboratorio se podría definir como aquella situación en la que el investigador produce las condiciones bajo las que va a observarse la conducta, con un absoluto control de las restantes variables. Aquí se entiende por control no sólo la acción directa y manipulativa del investigador sobre la variable independiente, sino sobre todas aquellas variables que, de una forma u otra, pueden llegar a alterar los resultados. De esta manera se puede conseguir una medida precisa y una observación exacta de cómo se comporta el sujeto frente una serie de estímulos.

Hemos dado del *experimento de laboratorio* quizás una definición dema-

siado rigurosa. Ya que, en buena lógica, el experimento de laboratorio exigiría del experimentador la utilización de una técnica muy perfeccionada. De momento, y sobre todo en el área de la conducta, no se cuenta con instrumentos precisos y eficaces como para asegurar un estricto control de las situaciones experimentales. No obstante, se puede pensar en la utilización de los experimentos de laboratorio siempre que se acepte la existencia de diversos grados de control y precisión.

Se han criticado los experimentos de laboratorio, como señalábamos anteriormente, por crear situaciones artificiales, y consiguientemente, porque no estudian la conducta en su dimensión real, lo cual limita enormemente la generalización de los resultados. Ahora bien, como muy acertadamente señala Festinger (1953), los experimentos de laboratorios no pretenden, ni han pretendido analizar el comportamiento en su complejidad real, sino "crear una situación en la cual se vean claramente cómo operan las variables en situaciones especialmente identificadas y definidas". El hecho de que las situaciones de laboratorio no sean reales, y que, por tanto, jamás las encontraremos en la vida real, no constituye una objeción grave a esta forma de trabajo. Ya que éstos están pensados, primordialmente, para comprobar el grado en que una determinada variable afecta la conducta en condiciones "altamente controladas".

Sin duda alguna, la planificación de un experimento de laboratorio implica una serie de graves dificultades que hemos de tener en cuenta. En primer lugar, la situación de laboratorio es poco propicia para que las variables manipuladas actúen con la misma intensidad con que actúan en la vida real. Por otra parte, otro de los graves inconvenientes que presentan los experimentos de laboratorio, es que en ellos es muy difícil estudiar la acción simultánea de varias variables. Debido a la precisión y al riguroso control que se consigue dentro del laboratorio, es casi imposible variar simultáneamente dos o más variables. Por esta razón, todas las conclusiones que obtengamos de los experimentos de laboratorio, deberán ser interpretadas en función de las grandes limitaciones que el método impone.

Debido a esta serie de graves dificultades, los experimentos de laboratorio tienen, como finalidad básica, la de probar aspectos o hipótesis muy concretos con respecto a relaciones entre factores muy específicos. Los experimentos de laboratorio no son, por tanto, muy sugerentes, ya que, debido a su riguroso control, sólo nos permite analizar, con mucho detalle, aspectos muy concretos de la conducta.

7. EXPERIMENTOS DE LABORATORIO "VERSUS" ESTUDIOS DE CAMPO

Las investigaciones de campo, y con esta etiqueta abarcamos todos los trabajos que se realizan en condiciones naturales, nos permiten estudiar la conducta dentro de un sistema abierto, en el que están presentes un conjunto

de variables interactuantes. Las características propias de estas investigaciones las constituyen la gran variedad de factores que deberemos afrontar y la falta de un control riguroso. Por dicha razón, las conclusiones que podamos obtener de tales investigaciones no siempre serán definitivas e inequívocas. No obstante, estos trabajos se constituyen en fuente constante de nuevas ideas e hipótesis. Estas hipótesis pueden ser adecuadamente probadas en experimentos de laboratorio, procedimiento idóneo para la consecución de unos resultados precisos.

En los experimentos de laboratorio se logra un absoluto control de todas las condiciones, así como una directa manipulación de la variable experimental. Como consecuencia de ello se puede llegar a obtener resultados concluyentes, siempre que las hipótesis sean lo suficientemente concretas y específicas. El experimento de laboratorio contrapone a la vaguedad e imprecisión de las investigaciones de campo, la precisión y relevancia de sus resultados.

Como contrapartida, la investigación de campo ofrece la posibilidad de estudiar la conducta en situaciones reales, como efecto de una multitud interactuante de variables. Sus resultados no serán, quizá, tan precisos, pero abrirán nuevos caminos a la investigación y suministrarán los elementos necesarios para la investigación de laboratorio.

De todo ello podemos concluir que estos dos procedimientos más que oponerse se complementan, y no puede concebirse el uno sin el otro. Frente a la amplitud de la investigación de campo nos encontramos con la limitación de los estudios de laboratorio que, a su vez, nos proporcionan el rigor y la precisión del que están desprovistos los primeros.

8. RESUMEN

Hemos intentado presentar la importancia de la observación en el método científico, vinculada, directamente, en una de sus principales fases: la recogida de datos. En la recogida de datos, que constituye el aporte de pruebas empíricas para la verificación de la hipótesis, la observación constituye el único procedimiento válido. Ahora bien, como hemos visto, no siempre la observación se ha aplicado con el mismo rigor y precisión. Esto nos ha llevado a considerar que si bien la observación es la base para la prueba empírica de la hipótesis, no siempre se aplica con el mismo rigor. De ahí, pues, en virtud de las limitaciones o restricciones que sometamos a la observación, en la fase de recogida de datos, aparecen una serie de métodos o sistemas que hemos denominado: *naturales*, *diferenciales* y *experimentales*. En la medida, pues, que las técnicas de observación utilicen procedimientos y técnicas que aseguren la validez de los datos, conseguiremos métodos mucho más precisos. Tenemos, pues, que la clave básica para la distinción de los métodos, en función de la observación, es el "control" que podemos tener de todos aquellos aspectos que intervienen en una situación concreta de estudio. A mayor

control, mayor rigor; a mayor control, también, el método es cada vez más experimental.

El valor de los métodos observacionales es función del grado de control que utilicemos en nuestras investigaciones. Por otra parte, como hemos visto, cada uno de estos métodos tienen sus ventajas e inconvenientes. Sería aventurado afirmar que el único método de trabajo es el experimental, ya que, en definitiva, toda ciencia aplica los diversos métodos de acuerdo a su nivel de desarrollo y exigencias prácticas.

9. BIBLIOGRAFÍA

- BEST, J. W.: *Cómo investigar en educación*. Morata, Madrid, 1972 (original inglés, 1959).
- BERNARD, L.: *Introduction à l'étude de la médecine expérimentale*. Flammarion, París, 1952 (existe trad. cast. en ed. Losada).
- BROWN, L. W., y GHISELLI, E. E.: *El método científico en psicología*. Paidós, Buenos Aires, 1959 (original inglés, 1955).
- CARTWRIGHT, D. P.: "Análisis del material cualitativo", en FESTINGER, L., y KATZ, D.: *Los métodos de investigación en las ciencias sociales*. Paidós, Buenos Aires, 1972 (original inglés, 1953).
- DAVIS, R. C.: "Methods of measuring and recording action", en ANDREWS, I. C. (ed.): *Methods of psychology*. J. Wiley, Nueva York, 1938 (existe traducción francés en P.U.F., 1952).
- FRAISE, P.: "La méthode expérimentale", en FRAISE, P., y PIAGET, J. (eds.): *Traité de psychologie expérimentale*, vol. I, P.U.F., París, 1970 (3.^a ed.).
- HAYMAN, R.: *The nature of psychological inquiry*, Prentice-Hall, Nueva Jersey, 1964.
- KERLINGER, F. N.: *Foundations of behavioral research*. Holt, Rinehart and Winston, Nueva York, 1964.
- PLUTCHIK, R.: *Foundations of experimental research*. Harper y Row, Nueva York, 1960.
- SHERIDAN, Ch. L.: *Fundamentals of experimental psychology*. Holt, Rinehart and Winston, Nueva York, 1971.
- WRICHT, H. F.: "Observational child study", en MUSSEN, P. H. (ed.): *Handbook of research methods development*. J. Wiley, Nueva York, 1960.

PRESENTACIÓN DEL MÉTODO EXPERIMENTAL

por RAMÓN BAYÉS

Me ha correspondido en este seminario la presentación del llamado método experimental. En realidad, el doctor Arnau ha centrado ya el problema por lo que, dado lo limitado del tiempo de que disponemos y, por contraste, la ambiciosa amplitud del tema a desarrollar, aunque quizá será inevitable caer en alguna repetición, trataré de prestar especial atención aquellos puntos que, por lo menos en el seno del seminario, puedan ser considerados como nuevos.

Recordemos que el doctor Arnau distingue entre experimentos de campo y experimentos de laboratorio, señalando que la palabra "experimento", tomada en sentido estricto, suele reservarse para estos últimos. Por mi parte y con el fin de evitar confusiones, debo indicar desde el principio que siempre que aluda al método experimental me referiré únicamente a investigaciones de laboratorio en las que, como ha señalado el doctor Arnau, el investigador ejerce un elevado control sobre las condiciones responsables de los fenómenos que nos interesan, si bien mi postura en cuanto a las posibilidades de dicho método para el análisis de las variables conductuales es, quizá, más optimista que la suya.

Antes de adentrarnos en el estudio del método experimental propiamente dicho, sin embargo, es posible que no sea superfluo recordar algo que si bien constituye la base de cualquier ciencia, no siempre se admite con facilidad cuando la materia de estudio es la conducta y el problema que este hecho plantea emerge de forma clara y explícita. Si no lo soslayamos es debido a que estamos íntimamente convencidos de su importancia.

Cualquier ciencia tiene, como punto de partida, la admisión de que el mundo se encuentra sujeto a un orden y que su particular objeto de estudio no sólo posee características totalmente físicas sino que se encuentra sometido a leyes estatales; la misión del investigador científico será descubrirlas. En nuestra opinión, si la Psicología desea ser una ciencia, no sólo debe *rechazar cualquier posible forma de dualismo "cuerpo-mente"* —el método científico sólo es aplicable a fenómenos físicos— sino que, guste o no, debe *admitir el postulado de determinismo*, es decir, que la conducta se encuentra sometida a leyes fijas y determinadas.

Si los fenómenos conductuales no se encontraran sujetos a leyes, un enfoque experimental de la Psicología tendría poca utilidad ya que si disponiendo las mismas condiciones y actuando bajo control adecuado no pudiera predecirse un resultado conductual en la replicación de un experimento, esta disciplina quedaría condenada, permanentemente, a coleccionar “experimentos” lo cual, en nuestra opinión, carecería de sentido.

Por nuestra parte, definimos a la Psicología como la *ciencia cuyo objeto de estudio son los fenómenos conductuales y la determinación de las condiciones materiales de su aparición*. En la medida en que consigamos averiguar estas condiciones podremos explicar, predecir y controlar la conducta. Creer que esto no es posible es sólo esto, una creencia; partir del supuesto de que esto es posible es un requisito necesario para su estudio científico; demostrar que esto es así ya no depende ni de las creencias ni de los supuestos sino de las pruebas empíricas, públicas y verificables, que podamos aportar. El grado de desarrollo de una ciencia no depende de sus teorías sino de sus frutos; por nuestra parte, no tenemos duda alguna de que si los supuestos de que partimos se demuestran correctos, la aplicación del conocimiento de las leyes conductuales al campo educativo, por ejemplo, nos proporcionará resultados sorprendentes. Debemos tener bien presentes, sin embargo, las palabras de Underwood (1957): “Rechazar el determinismo para toda o parte de la conducta humana equivale a rechazar la aplicación de las técnicas científicas de la misma”.

Hemos hablado de “fenómenos conductuales” y de “condiciones ambientales”; en otras palabras, de conducta y ambiente. Es conveniente que definamos nuestros términos. Desde nuestro punto de vista, consideramos que ambos deben ser contemplados en un sentido amplio. *Conducta* —palabra tan cargada, desgraciadamente de resonancias emotivas— es todo lo que un organismo *hace*. Andar, coger, salivar, etc., serán, por tanto, conductas; pero pensar y emocionarse también lo serán. Por tanto, cuando nos referimos a conducta no nos limitamos a acontecimientos externos sino que incluimos también los internos mientras tengan lugar en el mundo *físico*. *Ambiente* será todo aquello que puede afectar de algún modo a la conducta. La temperatura de la habitación en que nos encontramos, la sonrisa que nos dirige otra persona, la droga que nos administran, etc., serán otras tantas variables ambientales susceptibles de influir sobre nuestra conducta.

Para que el estudio de un fenómeno conductual concreto —externo o interno— pueda ser abordado científicamente debe cumplir, sin embargo, como cualquier objeto de estudio de cualquier ciencia, dos condiciones:

a) Tanto la conducta como las condiciones ambientales deben poder definirse de forma inambigua. De este importante tema les hablará la doctora Bartolomé.

b) Tenemos que disponer de técnicas o instrumentos para que el fenómeno pueda ser observado, registrado o medido de forma fiable, ya que los

datos obtenidos deben ser públicos y repetibles por cualquier investigador que posea los conocimientos o instrumental necesarios. Este tema ya ha sido desarrollado por el doctor Arnau.

El hecho de que en un momento determinado no podamos investigar determinados fenómenos conductuales —por no ser capaces de definirlos, observarlos, registrarlos o medirlos— no significa, por tanto, que dichos fenómenos no existan o que deban quedar para siempre al margen de un estudio científico, sino que la ciencia nada puede decir, en este particular momento, sobre ellos, y que las teorías o hipótesis que puedan ofrecerse, por muy convincentes y satisfactorias que intelectualmente puedan parecer, no formarán parte del edificio científico hasta que no puedan algún día, si es que pueden, ser validadas por hechos reproducibles bajo condiciones controladas.

Para que podamos hablar de leyes de conducta, deberemos ser capaces de mostrar la existencia de relaciones funcionales estables entre la conducta —nuestra variable dependiente— y otros hechos observables. El sistema de elección para conseguirlo es el experimento. La diferencia entre un experimento y otro sistema de investigación, tal como ya ha sido señalado, es, fundamentalmente, una diferencia de control. El experimento es la situación controlada por excelencia.

Nótese que —excepto en la cita textual de Underwood— hemos hablado de conducta y no de conducta humana. El hecho es plenamente intencionado. Muchos datos empíricos recogidos hasta el momento en el laboratorio en condiciones sumamente estrictas, indican que la conducta del hombre parece encontrarse sometida a las mismas leyes que rigen la conducta de las especies inferiores.

Aunque el hecho pueda repugnar a algunos, dista mucho, sin embargo, de ser revolucionario. Quizá sea instructivo recordar que las leyes de Mendel fueron descubiertas experimentando con guisantes y que, en Genética, algunas investigaciones suelen llevarse a cabo utilizando moscas. Al enfocar el tema de la información genética, Prevosti dice (1973): “Cuando explícitamente no digamos lo contrario, trataremos este tema refiriéndonos a los animales superiores, por el hecho de que entre ellos, biológicamente, se encuentra incluido el hombre. Conviene indicar, no obstante, que los aspectos fundamentales de lo que se dirá *son válidos para toda clase de sistemas vivientes*. La facultad de llevar, transmitir, compilar y actualizar la información, se basa en principios, mecanismos y utilización de materiales *comunes a todos los seres vivientes*” (los subrayados son nuestros).

Personalmente, creemos que el camino de los investigadores de la conducta pasa *necesariamente* por el laboratorio animal y que muy posiblemente deba recorrer etapas —las ha empezado a recorrer ya— parecidas a las ya cubiertas con pleno éxito por las ciencias médicas. Consideramos como un signo alentador —y no dudamos que significativo, en el contexto educativo en que nos encontramos— que, en 1971, se haya concedido a Skinner —inves-

tigador que ha pasado gran parte de su vida en el laboratorio animal y una de cuyas contribuciones científicas más importantes sea posiblemente el análisis de un cuarto de billón de respuestas proporcionadas por palomas, en condiciones experimentales— el Premio Kennedy para Investigación en el Retraso Mental. Su descubrimiento de que la conducta de los organismos —humanos o no— es función de sus consecuencias, equivale, como escribe Throne (1972) a que “la conducta del retrasado se encuentra limitada solamente por nuestro esfuerzo y habilidad para aplicar, de forma contingente, las infinitas consecuencias potencialmente disponibles para mejorarla”.

El hecho que acabamos de relatar nos ayuda a poner sobre el tapete otro tema polémico ya que no sólo comparamos la conducta de los animales inferiores con la conducta humana “normal” sino con la de seres humanos —los subnormales— que podemos considerar poseen características patológicas. En efecto, algunos datos, de los que ya se dispone, permiten suponer que tanto la conducta “normal” como la “patológica” obedece a las mismas leyes conductuales, lo cual no significa que los organismos no puedan mostrar topografías muy diferentes.

En el experimento tratamos de establecer la existencia de una relación funcional entre un aspecto de la conducta —variable dependiente— y un aspecto del ambiente —variable independiente— manteniendo neutralizadas las demás variables ambientales que también podrían influir, al mismo tiempo, sobre la primera y oscurecer la posible relación entre ambas.

El hecho de que una conducta determinada siga en el tiempo a una manipulación ambiental concreta no es suficiente para establecer una relación entre ambas. La garantía ideal la obtenemos, posiblemente, cuando manipulando la variable independiente de forma sistemática podemos predecir la aparición, mantenimiento o desaparición de la variable dependiente cuantas veces queramos y en cualquier organismo individual; tal grado de control, sin embargo, en muchas ocasiones, es difícil de obtener.

Podemos resumir las ventajas del método experimental, utilizando las palabras, ya clásicas, de Woodworth y Schlosberg (1954):

“1. El experimentador puede hacer que el fenómeno se produzca cuando lo desee; así, puede estar perfectamente *preparado* para observarlo con preservarlo con precisión.

2. Puede *repetir* sus observaciones en las mismas condiciones para su comprobación, y puede describir estas condiciones, dando así la posibilidad a otros experimentadores de repetirla y lograr una comprobación independiente de sus resultados.

3. Puede *variar las condiciones* de una manera sistemática y anotar las diferencias en sus resultados” (pág. 2).

Dadas las condiciones en que debe desarrollarse el experimento, la investigación, utilizando como sujetos animales inferiores, nos ofrece la siguientes ventajas:

- 1.º) *Elimina muchos problemas de tipo ético o jurídico.*
- 2.º) *Aumenta nuestro control sobre la historia conductual del organismo,* la cual pudo tener gran influencia en los resultados.
- 3.º) *Hace posible utilizar el mismo organismo durante largos períodos de tiempo,* en situaciones experimentales que pueden prolongarse, de forma continua —las 24 horas del día— incluso durante varios meses.
- 4.º) *Disminuye el riesgo de contaminación de los resultados* que posiblemente es máximo cuando el investigador y el sujeto experimental pertenecen a la misma especie.

Habrán muchos, de todas formas, que verán muy débil y lejana la relación existente entre, por ejemplo, una rata apretando una palanca y un niño de siete años aprendiendo matemáticas en el aula. Pueden venir en ayuda de nuestra esperanza en el método experimental, las imágenes de Arquímedes corriendo desnudo por las calles tras el baño, de Franklin levantando su cometa, de Fleming observando un cultivo de bacterias estropeado, de Becquerel manoseando una placa fotográfica improvisada o, quizá mejor, de Pavlov contando cuidadosamente las gotas procedentes de la salivación de un perro. Todos estos acontecimientos —y tantos otros— han tenido repercusiones importantes sobre las vidas de millones de seres humanos. Dado que, hasta el presente, no creo que podamos sentirnos especialmente orgullosos de los logros conseguidos por nuestras técnicas educativas, es posible que debamos contemplar con interés acrecentado los datos procedentes del laboratorio experimental y fomentar las investigaciones en este terreno ya que los mismos pueden ofrecernos una alternativa prometedora.

Existen muchas preguntas que podríamos hacernos: ¿hasta qué punto podemos generalizar los datos encontrados en un experimento? ¿Cómo llevar a cabo, en la investigación conductual, el grado de control a que hemos aludido? ¿Qué aspecto de la conducta puede ser adecuado para su estudio experimental? ¿Hacia dónde orientar la investigación conductual? ¿Cuáles son las leyes conductuales ya descubiertas en el laboratorio y que pueden tener repercusión en el campo educativo? ¿Cómo hacer descubrimientos?

Algunos de estos interrogantes —u otros muchos que podrían plantearse— serán posiblemente despejados en el transcurso del seminario, otros requerirán un curso; finalmente, hay también algunos para los que es posible que no exista ninguna respuesta en la actualidad.

Quisiéramos resumir nuestro punto de vista señalando que, si deseamos enfocar científicamente el estudio de la conducta, debemos seguir la misma senda que las demás ciencias. Como dice Skinner (1969): “El objetivo básico no es averiguar si la conducta es instintiva o aprendida, como si tales adjetivos describieran esencias, sino identificar correctamente tanto las variables responsables que la han originado como aquellas que normalmente la controlan... El objetivo importante es empírico: ¿cuáles son las variables relevantes?”

En mi opinión, sólo el método experimental puede permitirse contestar a esta pregunta.

REFERENCIAS

- PREVOSTI, A.: "Aproximació a problemes epistemològics des de la biologia actual". *Perspectiva Social*, 1973, 1 (2), 7-29.
- SKINNER, B. F.: *Contingencies of reinforcement*. Appleton-Century-Crofts, Nueva York, 1969.
- THRONE, J. M.; BEYOND; SKINNER, B. F.: *Mental retardation*, 1972, 10 (3), 44.
- UNDERWOOD, B. J.: *Psychological research*. Appleton-Century-Crofts, Nueva York, 1957.
- WOODSWORTH, R. S., y SCHLOSBERG, H.: *Experimental Psychology*. Holt, Rinehart and Winston, Nueva York, 1954. (Traducción: *Psicología experimental*. Eudeba, Buenos Aires, 1962.)

ESTUDIOS DE LAS VARIABLES EN LA EXPERIMENTACIÓN EDUCATIVA

por MARGARITA BARTOLOMÉ

CONSTRUCTOS Y VARIABLES

La terminología en la investigación experimental psicopedagógica se caracteriza por una relativa imprecisión y ambigüedad en su uso. De ahí la necesidad de definir en nuestro estudio el contenido de los términos utilizados.

En la elaboración de teorías que expliquen los mecanismos subyacentes a la conducta observada, los científicos suelen acudir a la creación e invención de *constructos*. Este término, por tanto, hace referencia a conceptos que han sido postulados con un propósito científico. A veces se los llama "constructos hipotéticos" para indicar que no se los considera como objetos y eventos reales. Mouly los define como "productos de la imaginación de científicos. Ellos son conceptos imaginarios que ofrecen una estructura alrededor de la cual puede ordenarse el pensamiento. En las primeras etapas del desarrollo científico, estos constructos eran extraídos generalmente de la experiencia personal. En etapas posteriores se postulan en términos de eventos que son inferidos más que directamente observados".¹

La mayoría de las teorías acerca de la conducta incluyen constructos tales como inteligencia, motivación, aprendizaje, etc. Hay que tener en cuenta que estas palabras pueden poseer una doble acepción. Por ejemplo a un nivel observacional nosotros podemos hablar de inteligencia, partiendo de las puntuaciones obtenidas en un test *X* e indicar qué niños son más inteligentes que otros. Sin embargo, como constructo científico, el término inteligencia puede usarse para explicar la variabilidad de los sujetos en la resolución de problemas a pesar de utilizar en el aprendizaje métodos similares.

En el desarrollo de constructos, Travers, siguiendo a Hull, señaló como condición esencial el evitar la *petición de principio*. Veamos un ejemplo: "En el estudio de la conducta de resolución de problemas, se solía explicar en

1. Mouly, G. J.: *The Sciences of Educational Research*, p. 59. Van Nostrand Reinhold Company. Nueva York, 1970.

el pasado tal conducta en función de un constructo llamado inteligencia. Este procedimiento implica un razonamiento circular, pues las conductas específicas de resolución de problemas se utilizan como base para postular una capacidad subyacente a la que se llama inteligencia, y luego, esa capacidad subyacente, se utiliza para explicar la conducta de resolución de problemas sobre la base de la cual se la había derivado originalmente. En tal situación, la invención de un constructo no sirve a ningún propósito útil".²

En realidad, buena parte de lo que se denomina constructo hipotético en psicología representa mecanismos hipotéticos que median entre el estímulo y la respuesta. Travers llega a afirmar: "Un constructo útil siempre vincula una condición antecedente con una condición subsiguiente".³ Sin embargo, en este tema, preferimos utilizar la terminología de Kerlinger que le otorga una más amplia significación.⁴

El término *variable* hace referencia a una propiedad o característica que puede adoptar distintos valores dentro de un grupo de ellos al que se denomina dominio de la variable. Estadísticamente podemos encontrar el modelo en la definición de una variable aleatoria. Es decir, si se tiene un acontecimiento cuyas posibilidades indicamos con

$$X_1, X_2 \dots X_n$$

donde las X_i son *valores numéricos* o *modalidades correspondientes a un atributo*. Además, sabemos que es atribuible a cada X_i una probabilidad p_i , de tal manera que

$$\sum_{i=1}^n p'_i = 1$$

diremos entonces que ha quedado definida una variable aleatoria, lo que suele expresarse con la relación:

$$\left\{ \begin{array}{l} X_1, X_2, \dots, X_n \\ p^1_1, p_2, \dots, p_n \end{array} \right. \quad ^5$$

2. Travers, R. M. W.: *Introducción a la investigación educacional*, p. 32. Paidós. Buenos Aires, 1971.

3. *Ibidem*, p. 32. Desde este punto de vista la identificación entre constructo hipotético y variable interviniente aparece clara. En una nota, el autor señala cómo P. E. Meehl y K. Mac Corquedale (1948) intentaron trazar una distinción entre estos dos conceptos, pero el distingo que propusieron suscitó objeciones. Al tratar el tema de las variables intervinientes explayaremos más este punto.

4. Kerlinger, F. N.: *Foundation of Behavioral Research*, pp. 31 y 33. Holt, Rinehart and Winston. Nueva York, 1965.

Para Kerlinger existen dos niveles en los que puede operar el científico: El nivel observacional de donde extraemos los datos que nos brinda la realidad empírica y el nivel teórico en el que elaboramos hipótesis a base de conceptos científicos denominados constructos. La relación entre uno y otro nivel, así como los modos para pasar del nivel en que se ha construido el constructo al nivel observacional, es de capital importancia en toda experimentación científica.

5. Toranzos, F.: *Teoría estadística y aplicaciones*, p. 28. Kapelus, Buenos Aires, 1971.

Una variable, por tanto, es un símbolo al que pueden serle asignados valores numéricos o modalidades diversas. Variables importantes en psicología y educación con el sexo, la inteligencia, la clase social, la motivación,⁶ De ellas, unas reflejan una variación cuantitativa como los diferentes C.I. obtenidos en un test de inteligencia; otras señalan una variación cualitativa como puede encontrarse al referirnos al sexo o a la clase social. En la investigación experimental, la *definición* correcta de las variables, la *función* que desempeñan dentro del diseño elaborado, así como el *control* de las mismas, de acuerdo con su naturaleza, representan tres problemas fundamentales que todo científico debe abordar.

Definición de variables. Al intentar planear una investigación la dificultad primera con que solemos encontrarnos reside en la definición correcta de los términos en los que está expresada la hipótesis. Ello implica indicar claramente lo que queremos controlar así como el tipo de relaciones establecidas entre estas variables. En el campo educativo la dificultad aumenta extraordinariamente por la complejidad de los procesos en los que intervienen múltiples variables y el esfuerzo (que raramente alcanza el éxito) para salvar la distancia entre lo que pretendemos medir y lo que realmente medimos. Cuando yo afirmo: "La madurez lectora influyó en la rapidez para el aprendizaje de la lectura"; debería preguntarme: "¿Qué entiendo por madurez lectora?" o "¿A qué tipo de aprendizaje lector me estoy refiriendo?..."

Existen dos modos para definir las variables con las que trabajamos. El primero sería utilizar palabras que expliquen en qué consiste el fenómeno a estudiar, es decir, utilizando otros constructos en la definición. El segundo asignaría significados a la variable especificando las actividades u "operaciones" que han de realizarse para medirla.

El primer tipo de definición se denominaría: *constitutiva* y el segundo *operativa* u *operacional*. Las diferencias entre ambos son considerables.

Una definición *constitutiva* tiene un objeto más ambicioso y rico. Pretende mostrarnos la esencia del fenómeno. Intentará, por tanto, presentárnoslo en sus características fundamentales. Muchas veces, en cambio, este tipo de definiciones se reserva a los constructos hipotéticos que, directamente, no son susceptible de un control riguroso.

En la medida que una variable ha sido definida constitutivamente, la generalización de la experimentación se vuelve más fácil y las conclusiones pueden transferirse a otros planteamientos científicos. Sin embargo, existe un riesgo inminente: la validez de esas mismas conclusiones. En efecto, ¿es realmente cierto que mi experimentación ha trabajado con las variables tal como han sido definidas?... Diríamos entonces que: mientras el plantea-

6. La relación de las variables con los constructos es evidente. Kerlinger afirma: "Algunos científicos suelen llamar variables a los constructos o propiedades para su estudio". *Obra cit.*, p. 32.

miento y la interpretación de resultados se desarrollan a un mismo nivel, el trabajo experimental se lleva a cabo a un nivel evidentemente distinto. De ahí que, tales tipos de definiciones, se hayan rechazado con frecuencia en la investigación experimental.

Las definiciones *operacionales* pueden ser de dos clases: medidas y experimentales. Las primeras *describen* cómo la variable va a ser medida. En el ejemplo anterior, podría definir la madurez lectora como las puntuaciones obtenidas en un test de madurez lectora de primer grado de Victorino Arroyo del Castillo. Una definición operativa experimental indicaría, con todos sus detalles, las manipulaciones de dicha variable que ha de realizar el investigador. Si la variable que deseamos definir son los métodos para el aprendizaje de la lectura,⁷ debería indicarse todas las etapas que se han de recorrer para llevar a cabo la instrucción con cada método.

Desde luego las definiciones operacionales son los instrumentos fundamentales en la investigación experimental. Como afirma Kerlinger, ésta no puede llevarse a cabo sin observaciones y las observaciones no pueden recogerse sin instrucciones que nos digan la manera de hacerlo. Las definiciones operacionales serían tales instrucciones.⁸

Claro que las variables así definidas limitan extraordinariamente la significación de los constructos. Sin embargo, el uso de las definiciones operacionales han contribuido a la precisión y desarrollo de ciencias que, como en el caso de las ciencias psicológicas y pedagógicas, poseían un vasto vocabulario poco científico.⁹

CLASIFICACIÓN DE LAS VARIABLES

El estudio de las variables implica, siempre, un proceso clasificatorio. Una vez definida una variable nos surgen estas otras preguntas: ¿Qué función se le asigna en el proceso experimental? ¿Cuál es su naturaleza, de acuerdo con los fenómenos a los cuales se vincula? ¿En qué tipo de escalas puede ser

7. Las investigaciones experimentales que pretenden controlar la eficacia de los métodos o técnicas empleados para el aprendizaje de un determinado contenido de instrucción son muy numerosas.

8. Mc. Guigan plantea la cuestión en términos tajantes: "Ahora podemos notar que el primer paso en la resolución de un problema es preguntarse si los términos empíricos fundamentales pueden ser definidos operacionalmente... Si para todos estos términos no puede realizarse dicha operación, debemos concluir que el problema es irresoluble y que la hipótesis es improbable". Ver: Mc Guigan, F. J.: *Psicología experimental. Enfoque metodológico*, p. 43. Trillas, México, 1971.

9. Ver por ejemplo el artículo de Skinner, B. que cita Kerlinger: "The Operational Analysis of Psychological terms" en Feigl, H. and Brodbeck, ed.: *Readings in the Philosophy of science*, p. 586. Appleton, Nueva York, 1953.

El operacionalismo como corriente se basa en el supuesto de que la definición adecuada de las variables con las que la Ciencia trabaja es un prerrequisito para su avance. El exclusivismo en las definiciones operacionales ha dado lugar a una dura polémica. Ver por ejemplo, Underwood, B.: *Psychological research*. Appleton, Nueva York, 1957

medida? Cada uno de estos interrogantes responde a un criterio de clasificación. Desgraciadamente, los científicos no utilizan siempre el mismo criterio ni la misma terminología. De ahí que las interferencias entre las descripciones de las variables hechas por distintos autores sea continua.¹⁰

Travers ayuda al esclarecimiento de estas clasificaciones introduciendo una división de acuerdo con unos criterios. Aunque con variantes, seguiremos a este autor en el establecimiento de los mismos.

A) En función de los fenómenos a los cuales se vincula

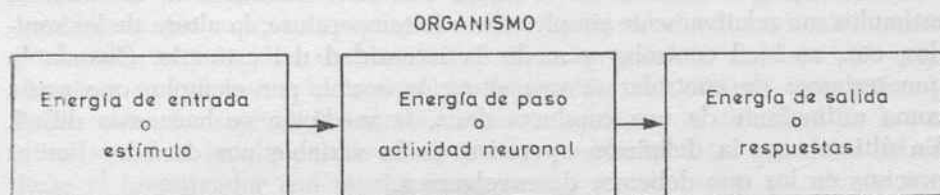
Esta clasificación es la más común en las investigaciones psicológicas.

A.1. Variables estímulo.

A.2. Variables respuesta.

A.3. Variables organísmicas.

Mc Guigan realiza una comparación entre el funcionamiento de una computadora y la dinámica de la conducta humana que abarca estos tres tipos de variables.¹¹



A.1) El término *estímulo* se incorporó a las Ciencias de la conducta proveniente de la fisiología, donde se le definía como un evento ambiental muy simple que producía una respuesta en un órgano receptor. Posteriormente los psicólogos tendieron a sustituir el término respuesta por el de experiencia, dándole entonces un sentido diferente, ya que por supuesto hay respuestas de las que el sujeto no es consciente. La insatisfacción en el empleo del concepto de experiencia dio lugar a una modificación en la definición de estímulo. Se le consideraba, entonces, como un evento o condición ambiental

10. Mc Guigan señala algunas de las numerosas clasificaciones de las variables que manejan los psicólogos: "Spencer (1948) considera variables de estímulo, orgánicas, de respuesta y de estado hipotético; Underwood (1957) discute variables del medio ambiente, de tareas, de instrucciones y variables de sujeto; Edwards (1968) usa la presente clasificación (de estímulo —energía de entrada—, organísmicas —energía de paso—, de respuesta —energía de salida—), de la misma manera que Woodworth y Schlosberg (1955). Este último agrega a la lista las variables antecedentes". *Obra. cit.*, p. 179.

11. Mc Guigan: *Obra cit.*, p. 178.

que producía un cambio sobre la conducta de forma demostrable. Así, una variable estímulo se transformó en lo que otros llaman una variable independiente, es decir, una condición que influye sobre la conducta siguiente.¹² Aun para quienes admiten la clasificación tripartita que estamos analizando, el término estímulo les resulta, a veces, excesivamente restringido y poco convincente para la experimentación psico-educativa actual. Intentan utilizarse para ello términos provenientes del campo de la ingeniería o de la informática. Por ejemplo: entradas de información o energía de entrada como veíamos en Mc Guigan. Así, los factores ambientales que influyen sobre la conducta han sido designados con diferentes nombres según el punto de vista de quien los usa. En el terreno educativo estas variables incluyen no sólo las características de los edificios, los manuales, las ayudas visuales y otros rasgos del ambiente físico, sino también la conducta de los maestros asesores, padres, directores y otras personas con las cuales entra en contacto quien intenta educarse.¹³

En muchos casos, la variable estímulo constituye una serie cualitativa. Por ejemplo, los experimentos relacionados con la presentación de material de forma visual o auditiva; o con los métodos empleados en el aprendizaje, etcétera.

En otros, la variable estímulo forma una serie cuantitativa. Cuando los estímulos son relativamente simples como la temperatura, la altura de los sonidos, etc., es fácil controlar y medir la intensidad del estímulo. Cuando lo que tratamos de controlar es una situación social, por ejemplo, que actúa como estimulante de una conducta dada, la medición se hace más difícil. En último caso la definición operativa de la variable nos dará los límites precisos en los que debemos desenvolvemos.

A.2) Las variables *respuesta* tienen una resonancia indudable en el terreno pedagógico. Los fines últimos de la educación se definen en función de las maneras deseables de respuesta a las situaciones que plantea la vida.¹⁴ Claro que estas respuestas se hallan en estrecha relación con cambios internos del sujeto y que son los que verdaderamente se buscan en la educación. La modificación y orientación de las actitudes radicales del hombre, su

12. La explicación de Travers que hemos sintetizado en estas líneas nos permite descubrir algunos orígenes de la confusión terminológica a la que aludíamos al principio. Ver Travers: *Obra cit.*, pp. 125 ss.

13. Travers: *Obra cit.*, p. 126. El lector interesado en el tema puede encontrar en este libro abundantes ejemplos de investigaciones educativas en las que se señalan las variables-estímulo utilizadas: índices de dificultad lectora, mediciones del monto de conducta de diversas categorías provenientes del maestro, empleo de ayudas audiovisuales y auditivas, etc. En cuanto a las investigaciones realizadas sobre la influencia del ambiente en la personalidad de los sujetos, en rendimiento académico o diversos aspectos de su educación pueden verse los libros de Illueca, L.: *Influjo del ambiente en la enseñanza media*. Dirección General de Enseñanza Media. Madrid, 1976; o en el de García Yagüe: *Condicionamientos ambientales de la personalidad*. Magisterio Español, Madrid, 1969, por citar autores españoles.

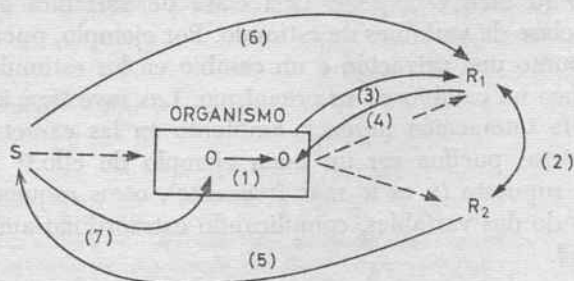
14. Travers: *Obra cit.*, p. 130.

sistema de valores, etc., han comenzado a ser objeto del estudio apasionado por parte de muchos investigadores en Ciencias de la educación, quedando en segundo término variables tan importantes como son el rendimiento escolar en determinadas materias, el aprendizaje de habilidades instrumentales (lectura, escritura, cálculo, etc.) o la ejecución en determinadas tareas escolares. Hay que señalar, sin embargo, que estos cambios internos a los que aludíamos antes, se expresan también, en cierta medida, a través de respuestas y es mediante el control de estas respuestas como llegamos a descubrirlos y comprenderlos.

Al igual que ocurría con los estímulos, las variables respuestas que se utilizan en la investigación pueden ser relativamente simples como los cambios respiratorios, o bastante complejos como el escribir a máquina, jugar al ajedrez, o adaptarse a la sociedad. Los instrumentos de medida que queramos utilizar deberán tener en cuenta el tipo de control de respuesta que deseamos escoger.

A.3) Las *variables organísmicas* se definen como los modos en los que los organismos pueden ser clasificados a partir de observaciones y medidas de sus características físicas, fisiológicas y psicológicas. La inteligencia, el sexo, el nivel o grado de desarrollo, etc., son variables organísmicas. Pero también lo son aquellas que haces referencia al nivel educacional intereses personales o capacidad de motivación. En una palabra, las variables organísmicas se refieren a todo tipo de diferencias individuales que podemos hallar entre los sujetos. La necesidad de su control y la dificultad de la medida (al menos en algunos de ellos) son dos problemas a los que debe enfrentarse el investigador con mucha frecuencia.

Veamos ahora las posibles relaciones que pueden darse entre las variables.¹⁵



1) $O_x \text{ — } O_y$ (o bien $O_y = f(O_x)$): *Problemas concernientes a la relación entre dos variables organísmicas.* Por ejemplo, la influencia del sexo en los

15. Mc Guigan: *Obra cit.*, p. 178.

cambios de altura de los alumnos, o en la madurez para el aprendizaje de la lectura a una edad determinada.

2) R_x — R_y (o bien $R_1 = f(R_2)$): *Problemas concernientes con las relaciones entre dos variables respuesta.* Por ejemplo, las respuestas a dos tipos de tests. Como en el caso anterior podemos indicar lo mismo de otro modo diciendo que:

Respuesta n.º 1 es función de la respuesta n.º 2.¹⁶

3) O_x — R_y (o bien $R_1 = f(O_x)$): *Problemas concernientes a las relaciones entre una variable organísmica y una respuesta.* Por ejemplo: ¿Difieren significativamente niñas y niños en la riqueza de vocabulario?

4) R_x — O_x (o bien $O_x = f(R_1)$): Es el caso inverso al problema anterior. Aquí se estudian las diferentes respuestas entre organismos y se intentan relacionar con una variable organísmica. Por ejemplo, predecir la inteligencia de los sujetos a partir de la habilidad para resolver puzzles.

5) R_x — S_y (o bien $S_1 = f(R_1)$): Este caso es muy poco frecuente, aunque puede darse. Edwards cita el ejemplo de la investigación de Cherman. En ella los niños han sido sujetos a estímulos dados y sus respuestas debidamente controladas. Si suponemos que cada estímulo provoca una respuesta específica, a través de las respuestas de los sujetos podemos predecir los estímulos. Es decir, *se intenta predecir la naturaleza del estímulo a partir de las variables de respuesta.*

6) S_x — R_y (o bien $R_1 = f(S_x)$): Éste es el caso más frecuentemente usado en las investigaciones psicológicas. *En estos problemas se estudian la relación entre la variable estímulo y la respuesta.* Por ejemplo, la eficacia de los textos programados sobre los tradicionales en el rendimiento en una determinada área de conocimientos.

En este tipo de experimentaciones puede variarse la intensidad del estímulo o bien, como en el ejemplo anterior, puede presentarse una variación cualitativa del mismo.

7) S_x — O (o bien $O = f(S)$): Una clase de variables organísmicas es función de una clase de variables de estímulo. Por ejemplo, puede interesarnos ver hasta qué punto una privación o un cambio en los estímulos ambientales del sujeto produce un cambio en su organismo. Las investigaciones realizadas para descubrir la interacción herencia-ambiente en las características personales de los sujetos pueden ser un buen ejemplo de ello.¹⁷

Existen, por supuesto (y es lo más frecuente), otros esquemas que ponen en relación más de dos variables, complicando extraordinariamente la posibilidad del control.

16. Mc Guigan utiliza esta última notación en tanto que Edwards lo hace de la forma primera. Ver Mc Guigan: *Obra cit.*, p. 197, y Edwards, A. L.: *Experimental design in Psychological Research*, p. 10. Rinehart, Nueva York, 1950. En esta clasificación hemos seguido preferentemente a este autor.

17. Véanse algunas de ellas en Anastasi, A.: *Psicología diferencial*. Aguilar, Madrid, 1966; o Meyer, W.: *La Psicología evolutiva y el proceso de la educación*. Troquel, Buenos Aires, 1968.

Las relaciones entre las diversas variables pueden estudiarse desde otro punto de vista. Madsen¹⁸ realiza una clasificación tripartita de las hipótesis a partir de las relaciones existentes entre dos tipos de variables: empíricas e hipotéticas.¹⁹

a) Aquellas que únicamente establecen relaciones entre variables empíricas (hipótesis empíricas puras).

b) Aquellas que establecen relaciones entre variables hipotéticas y empíricas (hipótesis parcialmente empíricas).

c) Las hipótesis que tratan exclusivamente de relaciones entre variables hipotéticas.

a) Dentro de las *hipótesis empíricamente puras* podemos incluir todas las que se han desarrollado en el apartado anterior y que pone en combinación las variables estímulo, orgánicas y de respuestas.²⁰

b) Las hipótesis parcialmente empíricas²¹ suponen que una de las dos variables, tiene un carácter hipotético.

$$1. H = f(E_{t-n})$$

$$2. H = f(E_{t-s})$$

$$3. R = f(H)$$

En este tipo de hipótesis se establece una relación entre un tipo de variable hipotética (H) con respecto a las condiciones estimulativas del ambiente medio (E), o bien, una relación entre la "conducta" (R) y una variable hipotética (3). La variable hipotética (H), puede ser función o bien de los estímulos previos (tanto internos como externos) (E_{t-n}), o bien función de los estímulos presentes (E_{t-s}). Por ejemplo: la necesidad de estimación como refuerzo de la conducta, etc.

c) *Hipótesis teóricas puras*. En ellas se formulan relaciones entre diversas variables hipotéticas, o bien su dependencia con respecto al factor tiempo. Según Spence estas hipótesis podrían ser representadas mediante la siguiente ecuación:

$$1. H = f(\text{tiempo})$$

$$2. H = f(H)$$

18. El problema, como hemos insinuado antes, es la dificultad de medir las variables orgánicas, ya que utilizamos para ello muchas veces las respuestas del sujeto.

De ahí que el esquema se nos quede en: $X_x \text{---} O_y \text{---} R_y$.

18. Madsen, K. B.: *Teorías de la motivación*, pp. 41 ss. Paidós, Buenos Aires, 1967.

19. Esta terminología puede referirse parcialmente al tipo de definición con que han sido tratadas (operacionales o constitutivas). Las variables hipotéticas pueden considerarse constructos no definidos operacionalmente.

20. Spence (1948) había indicado, únicamente, cuatro tipos de hipótesis: $R = f(R)$; $R = f(E)$; $R = f(O)$ y $O = f(E)$.

21. En la síntesis de las teorías de Spence y Madsen hemos seguido a Arnau, J.: *Fundamentos de Psicología experimental*. Ap. Policop, Barcelona.

Como ejemplo pueden verse los estudios existentes sobre la relación hábito-impulso de Hull y otros. Ambas variables son hipotéticas.²²

B) Clasificación de las variables en función del proceso experimental

En relación con el papel que desempeñan en el proceso experimental podemos clasificar las variables en:

- B.1. Variable independiente.
- B.2. Variable dependiente.
- B.3. Variables intervinientes.

B.1) La variable independiente

Su definición ha sido realizada muchas veces desde el ángulo de la relación causa-efecto. La variable independiente se consideraría *causa* de la variabilidad o modificación observada en la variable dependiente.²³ En la mayoría de las investigaciones pedagógicas el experimentador hace variar alguna situación del ambiente del niño o intenta, a continuación, observar el efecto en el rendimiento escolar o en algún otro aspecto de su conducta.²⁴

Junto al concepto de causación aparece *la ley de la variable única*: En cada experimento debe variar una circunstancia cada vez y mantener las demás fijas.²⁵ En el campo educativo la complejidad de las situaciones ha puesto en tela de juicio últimamente esta ley.

“Aparte de las limitaciones prácticas de la experimentación basada en la ley de la variable única, la objeción más fuerte que se le ha impugnado es la de hacer el experimento ya que establece una situación artificial que proporciona unos resultados no significativos.”²⁶

22. Un problema que nos plantea el trabajo con variables hipotéticas es el tipo de validación de las hipótesis, ya que la validación directa no es posible por su misma definición. Para ello se debe contrastar las consecuencias que implica la hipótesis a nivel observacional con lo que realmente se observa. En este caso, el razonamiento experimental, será del tipo hipotético-reductivo.

El esquema de Kerlinger sobre las relaciones entre variables definidas operacionalmente y constitutivamente y los diversos niveles de investigación establecidos, pueden ayudarnos a comprender las sub-hipótesis que debemos formular cuando debemos utilizar la validación indirecta. Ver Kerlinger: *Obra cit.*, pp. 36 y 37.

23. Kerlinger, la define “como supuesta causa de la variable dependiente que es el supuesto efecto”. *Obra cit.*, p. 39.

24. El análisis de los cánones de Stuart Mill ha influido en el concepto de causación hasta formar parte de la investigación experimental.

25. Withney: *Elementos de investigación*, Omega.

26. Mouly: *Obra cit.*, p. 320.

Más actuales son los conceptos de *causación múltiple* y *concomitancia*. Desde este punto de vista se considera que la experimentación puede operar en el contexto de una compleja interacción multivariada que caracteriza de manera más real la investigación educativa.²⁷

Un segundo y más amplio significado de la variable independiente procede del campo de la estadística. En muchas investigaciones científicas las variables aparecen en un continuo temporal, considerándose aquéllas que preceden la aparición de otros antecedentes necesarios de estas últimas.

Se acostumbra, así, a denominar la variable a predecir a variable dependiente (ya que depende de condiciones previas existentes denominadas independientes). El modelo estadístico que responde a este planteamiento es la ecuación de regresión

$$\hat{Y} = a + bx$$

donde y es la variable dependiente y x la independiente. Su expresión gráfica la encontramos en la recta de regresión.

Por supuesto, no es la naturaleza de las variables (como vimos en la clasificación anterior) las que les hace dependiente o independiente, sino la manera cómo se las utiliza.

Pueden existir variables independientes orgánicas, estímulo y, aunque menos utilizadas, respuestas. Hay que tener en cuenta que, sea cual sea la investigación que pretendamos hacer en la conducta del individuo, influirán tanto las variables ambientales estimulantes como las variables orgánicas. La labor fundamental consistirá no sólo en el control y manipulación de la variable independiente elegida, sino en mantener constante la influencia de las otras.

B.2) La variable dependiente

De alguna manera ya ha sido definida al hablar de la variable independiente. En la relación *causa-efecto* representa a este último. En la ecuación de regresión constituye el criterio que ha de predecirse a partir de otra u otras variables. Aunque, como vimos en el esquema general de relaciones entre variables, la variable dependiente puede ser de cualquier tipo, es más común que deseemos observar el efecto producido por un estímulo o variable orgánica en la conducta de un sujeto. De ahí que las variables respuestas sean las generalmente utilizadas en este tipo de variables. Un problema fundamental lo constituye la elección de la variable dependiente en una investigación. Sin duda, la influencia que una variable ejerce sobre la conducta de los sujetos puede ser múltiple y afectar a gran cantidad de respuestas.

27. Ésta es la base de los diseños factoriales complejos.

¿Cuál de ellas debemos considerar como más profundamente significativa para nuestra investigación? Por ejemplo, en las investigaciones sobre la eficacia del aprendizaje programado frente al tradicional, en diversas áreas de la instrucción, se midió durante bastante tiempo como variable dependiente el rendimiento alcanzado en dichas áreas por los sujetos (cantidad de respuestas acertadas en un test de instrucción). En la actualidad se estudia también el cambio de actitudes favorables, o no, hacia el aprendizaje provocado en los sujetos por el empleo de las técnicas programadas.²⁸

Algunas veces, el estudio de la correlación entre dos variables dependientes, puede indicarnos la conveniencia, o no, de eliminar una de ellas. Por ejemplo, si la correlación fuera de 0,95 ello nos indicaría que, prácticamente, estamos cuantificando la misma cosa y conviene suprimir una medida.

Podemos, en una investigación, estar interesados no únicamente por una medida terminal de la variable dependiente, sino por averiguar la influencia de cierto evento sobre una variable que puede cambiar a través del tiempo. Así, cuando tratamos de medir la influencia de la fatiga en el trabajo escolar, puede interesarnos constatarla en un momento dado, pero a veces se desea conocer cómo varía el ritmo y la calidad del trabajo al ir aumentando el grado de cansancio en el sujeto. Las curvas sobre el trabajo pueden ser un buen ejemplo de ello. Este tipo de *medida de desarrollo* forma un amplio capítulo en la investigación psico-pedagógica.²⁹ Finalmente podemos estudiar la posible retención del efecto experimental, a partir de *mediciones diferidas* de la variación dependiente.³⁰

B.3) Variables intervinientes

El confusionismo en la utilización de este término nos obliga a aclarar previamente el sentido que le damos en esta investigación.

En el proceso de una experimentación, junto a las variables dependiente e independiente, existe un buen número de ellas, bien procedente del medio

28. En el departamento de Ciencias Experimentales y Diferenciales de la Educación, Universidad de Barcelona, se han elaborado y experimentado unas cuarenta programaciones, aplicándole la mayoría de ellas a muestras de la población de Barcelona. De niveles sumamente heterogéneos (desde párvulos hasta jóvenes aprendices) se han utilizado dos tipos de programa: el lineal y el ramificado. Las áreas objeto de dichas programaciones han sido: Matemática, Aplicación a las industrias, Geografía e Historia, Lógica, Ortografía y Vocabulario, Ciencias Naturales, Arte e Información Convivencial. Mientras que, en general, se advierte una mejora en el aprendizaje del grupo que utiliza técnicas programadas, los cuestionarios de actitudes presentan impresiones muy diversas hacia aquéllas.

29. Cada vez más los estudios longitudinales van adquiriendo mayor volumen en el terreno educativo, introduciéndose en el campo de la didáctica experimental. Véase: Bartolomé, M.: *Tecnología didáctica de las matemáticas actuales*. Tesis Doctoral, Barcelona, 1969.

30. Ello es muy importante en el terreno educativo, ya que la eficacia de cualquier estímulo educativo se mide no únicamente por el efecto que produce sino por la *permanencia* del efecto. En las investigaciones sobre enseñanza programada siempre se aplica un test diferido para comprobar la retención del aprendizaje.



que rodea los sujetos experimentales, bien procedentes de las diferencias individuales de los mismos (variables orgánicas), que influyen y pueden desvirtuar los resultados. Todas estas variables extrañas al experimento, pero que pueden ejercer una influencia sobre él, se denominan variables intervinientes. Pero también existe otra posibilidad: que la experimentación en sí, al producirse, genere nuevas variables que, actuando en el organismo de los sujetos, dé lugar a una variabilidad en las respuestas. El esquema de Hull sobre el aprendizaje es un ejemplo claro de este segundo concepto de variable interviniente. Mientras que para el primer grupo de variables existen diversas posibilidades de control, para el segundo tipo se desconoce muchas veces los modos de poder hacerlo. La dificultad de poder definir estas variables operacionalmente, ha dado lugar a que se les identifiquen con los constructos hipotéticos a los que hacíamos referencia al comienzo de este estudio. Como ya entonces se indicó, esta identificación no es correcta. El término constructo es más amplio que el de variable interviniente. Por otra parte, este último término vemos que escapa a la consideración restringida de Tolman, Hull, Spencer o Mawrer y, en general, todos los autores conductistas y neoconductistas pretenden darle.

C) Clasificación desde el punto de vista de sus propiedades matemáticas

Una última clasificación de variables podría realizarse a partir de las escalas que se utilizan en su medición. La clasificación aquí ofrecida es de Stevens (1946).³¹

Entendemos por escala una regla o conjunto de reglas para la asignación de números a aspectos o relaciones empíricas. Se justifica su utilización sólo

31. De alguna manera ya ha aparecido esta clasificación al referirnos a los distintos tipos de variables: cuantitativas y cualitativas.



en tanto exista cierto isomorfismo entre las operaciones que podemos realizar con números y las que podemos realizar con objetos empíricos.

Veamos, en síntesis, estas escalas:

Las escalas nominales se utilizan en la medición de variables cualitativas. La escala ordinal al indicarnos el orden en que están colocados los eventos, aunque sin precisar exactamente la igualdad de distancias entre uno y otro, es la que más se ajusta al carácter de los fenómenos educativos. La estadística no paramétrica nos proporciona, actualmente, posibilidades para actuar a este nivel de medición.

También las escalas de intervalo se utilizan con mucha frecuencia en la investigación educativa. La mayoría de los tests estandarizados de rendimiento se construyen sobre la base de supuestos que, si se aceptan, llevan a considerar los puntajes como representativos de escala de intervalo. Sin embargo, hay que reconocer que no es tan sencillo probar la igualdad de diferencias entre los valores de la escala.

Las escalas de razón, por la exigencia de cero absoluto, son muy poco utilizables en el terreno de la Pedagogía ya que la mayoría de las variables utilizadas en ella no lo presuponen.

Veamos la clasificación de Stevens:³²

Escalas	Operaciones empíricas básicas	Estructura matemática del grupo	Estadísticas aplicables	Ejemplos
Nominal	Determinación de igualdad	Grupo permutativo $x' = fx$ fx significa cualquier sustitución biunívoca	Frecuencia moda Coef. de contingencia = correlación	Asignación de número a individuos y clases Asignación de modas y frecuencias a clases
Ordinal	Determinación de mayor y menor	Grupo isotónico $x' = fx$ fx significa cualquier función monotónica creciente	Mediana Percentiles Correlación ordinal	Escala de Mohs-Dmeza Orden de méritos Agrado de olores
De intervalos	Determinación de igualdad de intervalos o diferencias (unidad)	Grupo de función lineal $x' = ax + b$	Media Sigma Correlación	Temperatura Calendario Puntuaciones típicas
De razones	Determinación de igualdad de razones (cero absoluto)	Grupo de similitud $a' = ax$	Media geométrica Coeficiente de variación Transformación de decibelios	Longitud Escala de tonos (wels) Escala de intensidad (sones)

32. Stevens, S. S.: "On the Theory of Scales of Measurement", *Science*, CIII, pp. 677,680, 1946.

Coombs (1964) introduce algunas escalas intermedias, como las escalas *parcialmente ordenadas*, para abarcar ciertos instrumentos psicométricos de precisión que se encuentran entre las escalas nominales y las ordenadas y las escalas *métricas ordenadas*, que deben considerarse como una clase situada entre las escalas ordinales y las de intervalo.

CONTROL DE LAS VARIABLES

El control de las variables constituye la esencia del método experimental. Sin control es imposible la evaluación precisa de los efectos de una variable independiente sobre otra dependiente.

El propósito del control es asegurar las condiciones necesarias para poder averiguar este efecto, indicando lo que se debe a él y lo que se debe a otros factores. En el campo de la educación esto no es tan sencillo.

Campbell y Stanley, al presentar los diseños experimentales, señalan múltiples factores que comprometen la validez interna y externa del experimento.³³

Las fuentes de variación proceden bien de las diferencias individuales de los sujetos (ejemplo: inteligencia, grado de madurez, sexo, etc...), bien de las características diversas situacionales (ejemplo: el ambiente ruidoso, clima, etc...) y finalmente aquellas que proceden del sistema de la medida utilizada, instrumentos e incluso de las actitudes, expectativas y características personales del propio investigador.³⁴

Resumiendo los componentes fundamentales de la varianza que encontramos en la medida de la variable dependiente, vemos que:

VARIANZA	$\left\{ \begin{array}{l} \text{sistemática} \\ \\ \text{de error} \end{array} \right.$	$\left\{ \begin{array}{l} \text{primaria: producida por la variable independiente} \\ \\ \text{secundaria: producida por variables extrañas orgánicas o situacionales} \end{array} \right.$
		$\left\{ \begin{array}{l} \text{diferencias individuales no controlables} \\ \\ \text{errores de medición} \end{array} \right.$

33. Respecto a la *validez interna* estos factores son: historia, maduración, los efectos de un test al pasarlo por segunda vez, instrumentos, regresión estadística, mortalidad experimental e interacción de la selección-maduración, etc... En cuanto a la *validez externa* encontramos como factores influyentes: el efecto reactivo o interacción de la prueba; los efectos de interacción de la selección sesgada y la variable experimental; la interferencia producida por múltiples tratamientos. Los diseños experimentales deben señalar el grado de control ejercido sobre estos factores. Ver Campbell: *Experimental and quasi-experimental designs for research*. Rand M. C. Nally, Chicago, 1963.

34. La importancia, cada vez mayor, que se otorga al investigador como posible fuente adicional de varianza en los datos obtenidos, ha llevado a algunos autores a otorgarle un puesto paralelo al de las variables situacionales, sin incluirlas, propiamente, en estas últimas,

Tres serán los objetivos del investigador:

- disminuir en la medida de lo posible la varianza atribuible al error y cuando no es posible controlar su magnitud
- controlar el efecto de la variable independiente procurando que éste aparezca con toda claridad, mediante una manipulación o selección adecuada de los valores
- controlar las variables extrañas que intervienen en la experimentación y que proceden bien de los propios sujetos, bien de las situaciones o bien del propio investigador. Un control adecuado de estas variables influirá, sin duda, en el control de la variable independiente y aumentará, con ellos, la validez del experimento.³⁵

Veamos ahora algunos modos prácticos de controlar estas variables.

a) *Control de las variables extrañas*

Hay que tener en cuenta que en una experimentación educativa innumerables variables extrañas a la investigación, pero de alguna manera relacionadas con las variables que estamos controlando, pueden aparecer.³⁶

En primer lugar necesitamos una *especificación* de las variables extrañas a controlar. Desde luego que en el caso de investigaciones educativas y cuando se trabaja sobre todo en los propios centros escolares o escogemos el nuestro por conglomerados, algunas de ellas deben tenerse especialmente en cuenta.

Una lista no exhaustiva de ellas podría ser:

a) Características de los alumnos

- inteligencia general
- edad cronológica
- ejecución previa en el campo de la experimentación
- hábitos de estudio
- personalidad (actitudes, ideales, intereses)
- condiciones físicas
- sexo
- raza
-

b) Factores educativos

Factores de enseñanza

1. técnicas instructivas
2. procedimientos de manejo de clases
3. situación del maestro con respecto al factor experimental

35. Mc Guigan se refiere a dos tipos fundamentales de control: de la variable independiente y de las variables extrañas. Ver Mc Guigan: *Obra cit.*, p. 149.

36. Recordamos el apartado que hace referencia a las variables intervinientes.

Factores generales escolares

- tiempo empleado en las actividades escolares
- organización escolar
- características de la clase como grupo

Factores extraescolares

- participación de las actividades extraescolares
- vida familiar del alumno
- medio social y cultural
- intereses de la comunidad y actitudes ante la escuela

c) Factores ambientales inmediatos a la investigación

- ambiente del local
- hora de la investigación en relación con el conjunto de las actividades
- presencia o ausencia del investigador y su actuación en el experimento...

La tarea a realizar consiste en determinar cuáles de éstas u otras variables pueden influir en la investigación. Algunas de ellas podrán ser controladas y otras no, pasando su variabilidad a sumarse a la del error producido por las múltiples diferencias individuales. Sin embargo, si la influencia es muy grande, el investigador debería preguntarse si no debe abandonar el experimento. Muchas veces una actitud honrada obligaría a suspender bastantes investigaciones o a presentarlas como simples estudios exploratorios.

Para una mejor aclaración de los diferentes tipos de control podemos ver, ahora, las técnicas que intentan controlar la variabilidad entre los sujetos y las que controlan las diferencias situacionales.

b) *Control de las diferencias entre los sujetos*

Selección al azar. Cuando pretendemos buscar el efecto de la variable independiente sobre la dependiente, podemos dividir los sujetos seleccionados en dos o más grupos que son tratados de modo distinto. La asignación de sujetos a grupos debe realizarse de modo que las características de aquéllos no se interfieran significativamente con los resultados de la experimentación. El método más aceptado en el terreno científico es la asignación al azar a los grupos experimentales de modo que podamos creer —dentro de los límites del error del muestreo— que todas las características asociadas a ellos han sido randomizadas. En este caso, poseemos pruebas estadísticas de significación para ayudarnos a decidir cuándo las diferencias observadas entre los grupos son suficientemente mayores que las que podrían esperarse del simple azar. Teóricamente, este método es el único posible para controlar todas las variables extrañas posibles y puede utilizarse cuando el investigador no le interesa conocer cuáles son estas variables.

Grupos equivalentes. Cuando conocemos la influencia decisiva de alguna variable con respecto a la investigación (por ejemplo: el C.I.), podemos intentar que los grupos formados puedan igualarse con respecto a esa variable.

Estadísticamente podemos afirmar su equivalencia cuando no difieren significativamente ni su media ni su varianza. Por supuesto, el problema se agudiza cuando debemos hacer equivalentes los grupos utilizados más de una variable como criterio.

Existen varios modos prácticos de hacerlo:

— sistema de pares: se buscan dos sujetos con el mismo valor en la variable que deseamos controlar y se sitúa cada uno en un grupo distinto, escogiendo tantos pares como número de sujetos deseamos que posea la muestra.

Una forma natural de apareamiento es el realizar los grupos a partir de parejas de gemelos que suponemos relativamente similares en sus peculiaridades físicas y en algunas psíquicas.³⁷

c) Sección homogénea

Cuando es factible, podemos controlar la variable procurando *escoger a todos los sujetos que se ven influidos similarmente* por ella. De esta forma, si suponemos que el sexo puede influir en la investigación, escogeremos a todos los individuos del mismo sexo. Esta selección dificulta el proceso de generalización ya que debemos cuidar de no extender nuestros resultados a los sujetos del sexo contrario (siguiendo el ejemplo anterior).³⁸

d) Análisis de covarianza

Existe una forma de control que utiliza el método estadístico: es el análisis de covarianza. Puede ser usado cuando no nos es posible la elección previa de los sujetos al azar o utilizando procedimientos de equivalencia. En las investigaciones educativas, cuando debemos utilizar grupos de clases intactos, puede ayudarnos poderosamente en el control de las diferencias entre sujetos. Su fundamento es el siguiente:³⁹

Si hemos obtenido en la medición de la variable dependiente en varios grupos, una diferencia significativa entre las medias de los mismos, podemos estar tentados de creer que dicha diferencia se debe al efecto de la variable independiente. Pero si tenemos razones suficientes para pensar que los grupos

37. Las investigaciones deben realizarse dentro de un amplio margen de reserva ya que existen suficientes pruebas que no permiten afirmar la similitud necesaria en aptitudes, ni siquiera en gemelos univitelinos. La acción poderosa del ambiente, incluso antes del nacimiento, nos impide guiarnos únicamente por la clave genética de los homocigóticos.

38. Como puede verse, al tratar del control de las variables orgánicas sólo podemos utilizar dos tipos de control: de selección y estadístico. El tercero que corresponde a la manipulación física de la variable (si estamos clasificando el control desde el punto de vista de la posible manipulación por el investigador) se lleva a cabo con la variable independiente generalmente.

39. Lewis, D. G.: *Experimental designs in Education*, p. 172. University, London Press, Londres, 1968.

diferían con respecto a alguna variable relacionada con la investigación, sin duda esta última ha debido influir modificando la verdadera significación de los promedios obtenidos. En el análisis de covarianza, nosotros, estudiamos primero la conexión entre la variable dependiente Y y la variable extraña X dentro de cada grupo a través de una ecuación de regresión. Esta regresión puede diferir o no de grupo a grupo, es decir, la diferencia *entre grupos* puede, o no, ser significativa, lo que debería calcularse. En el caso último inferíamos una regresión de la población común a todos los grupos. Esto podría ser estimado mediante un promedio de todas las regresiones separados dentro de los grupos. Una vez que comprobamos la no significación de la diferencia entre las regresiones obtenidas en los distintos grupos, procedemos al "ajuste" de las medias de Y a partir de las diferencias en X .

Estas medias ajustadas se hallan a partir de la ecuación de regresión promedio. Un último paso consiste en averiguar el grado de significación entre las medias ajustadas.

Mc Nemar propone este ejemplo sencillo para clasificar el proceso: Sospechamos la influencia que puede tener la inteligencia sobre las puntuaciones obtenidas en un test de rendimiento. Los promedios en el primer test han sido de 60 para el grupo A y 70 para el grupo B.

El coeficiente medio en el test de inteligencia ha sido 105 para el grupo A y 111 para el B. Ahora bien, si existe una apreciable relación entre estas dos variables⁴⁰ los resultados podrían interpretarse a partir de las diferencias en inteligencia de los dos grupos. Parece pues lógico usar la ecuación de regresión para estimar las puntuaciones de rendimiento a partir de la inteligencia, como una base para predecir hasta qué punto las diferencias en rendimiento pueden aumentar a causa de las diferencias entre los grupos en el C.I. Efectuadas las ecuaciones de regresión, podemos predecir, a partir de los C.I. medios de inteligencia, una puntuación media en rendimiento de 62 para el grupo A y 62 para el grupo B. La diferencia en nuestra predicción es de 6 puntos; por lo que podríamos afirmar que 6 de los 10 puntos obtenidos al establecer la diferencia entre las medias encontradas en el test de rendimiento de los grupos A y B, pueden atribuirse a la falta de comparabilidad entre los grupos en inteligencia. La nueva cuestión que se nos plantearía sería el conocimiento del error muestral propio para usar al evaluar la significación de la diferencia ajustada. La búsqueda de estas medias "ajustadas", y de la prueba estadística que permita controlar la significación de sus diferencias, constituye la esencia del análisis de covarianza.⁴¹

40. Mc Nemar, G.: "Psychological Statistics"; en *Obra cit.*, p. 362. Por supuesto, antes de proceder al ajuste de medias necesitamos justificar que las regresiones obtenidas en los dos grupos proceden de una misma población, y, por tanto, que podemos utilizar un índice de regresión común.

41. Para no alargar este trabajo renunciamos a describir el procedimiento estadístico que puede encontrarse en cualquiera de los dos autores anteriormente citados.

e) *Utilizando a los sujetos como su propio control*

Un modo de proceder a la investigación, que permite eliminar las variables que surgen a partir de las diferencias posibles entre grupos de sujetos, es utilizar el mismo grupo para controlar el efecto de la variable independiente. Este método lo encontramos en el diseño clásico de grupo único de Mc Call.

$$[T_i - O - T_f - C_1] [T_i - F_E - T_f - C_2]$$

Hay bastantes casos en los que este diseño no puede utilizarse, en especial cuando la influencia de una condición experimental hace imposible el usar a los mismos sujetos para otra condición experimental.

En la práctica la mayor dificultad es conseguir encontrar las mismas condiciones para el test inicial. En el terreno pedagógico, si se trata de ver el efecto de un tipo de tareas sobre una determinada variable, las unidades de trabajo han de ser:

- de la misma dificultad
- igualmente interesantes
- otorgando la misma cantidad de tiempo para el desarrollo de cada una, o bien si el grupo es control, la misma cantidad de tiempo entre el test inicial y el final.

Puede usarse si el cambio producido por los factores perturbadores es insignificante o puede controlarse; si el cambio producido en la situación por el factor experimental no está condicionado por ningún factor anterior y si el cambio producido por la acción de cada factor es mensurable en unidades iguales.

CONTROL DE LAS VARIABLES SITUACIONALES

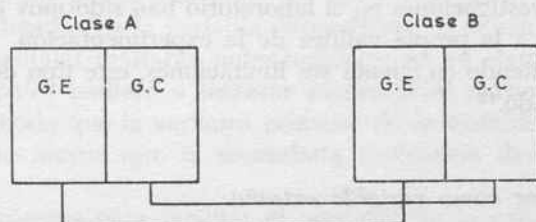
Además de realizar el control de las variables interindividuales, es necesario ejercerlo sobre las variables situacionales extrañas al experimento pero que pueden intervenir, en él, modificando los resultados. Esta nueva fuente secundaria puede ser de excepcional importancia en el recinto escolar: el tipo de escuela, las características de los maestros, el medio ambiente social, las distintas técnicas didácticas utilizadas en el aprendizaje, etc... constituyen una formidable fuente de variabilidad que debe controlarse.

Hay métodos para llevar a cabo este control:

1. Mantenerlas constantes
2. Randomizarlas
3. Manipularlas sistemáticamente y separadamente de la variable independiente
4. Eliminarlas mediante una restricción de las condiciones

1. *Mantener las condiciones constantes.* Se intenta que éstas sean similares en todos los sujetos que participan de la investigación: se han de utilizar los mismos tests, siguiendo el mismo orden en la ejecución, aplicándoles la misma persona, siguiendo unas pautas determinadas, etc... Desgraciadamente, esto, no puede realizarse en todas las variables extrañas que intervienen en el experimento, por lo que se ha de recurrir a otros procedimientos.

2. *Randomizar las variables situacionales.* El experimentador, en caso de una imposibilidad de mantener las variables situacionales fijas, intenta a veces randomizar o balancear ciertas variables situacionales. Por ejemplo: si no es posible el mismo maestro en ambos tratamientos, el experimentador puede dividir cada clase en dos grupos y asignar al azar los sujetos que corresponden a cada mitad. Una de estas partes correspondería al grupo control y otra al experimental.



3. *Manipulación sistemática de las variables extrañas.* En esta manipulación sistemática tenemos dos técnicas fundamentales:

- a) el contrabalanceo
- b) el diseño factorial

a) Este primer procedimiento intenta controlar los efectos de orden. En muchas investigaciones educativas es necesario usar una secuencia de condiciones experimentales, y de control, en orden a constatar los efectos progresivos. Por ejemplo: nos interesa conocer la relación entre distintos tipos de práctica y la fatiga, para evitar que el orden de presentación de los tipos de práctica (suponemos dos) influya en los resultados, podemos otorgar al azar a cada sujeto el orden por el que realizará su experimento. El contrabalanceo permite, sin embargo, simplificar la técnica anterior dividiendo al grupo en dos mitades; la primera realizará primero la práctica A y luego la B y la segunda mitad lo hará en sentido inverso (primero la B y luego la A).

Al extender el principio de contrabalanceo a un gran número de variables, el número de ordenaciones (y, por tanto, de sujetos) aumenta tanto que se vuelve impracticable. De ahí que se utilice el contrabalanceo incompleto.

Mc Guigan señala una dificultad que puede hallarse en el contrabalanceo: "Al emplear esta técnica, se supone que el efecto que produce al presentar una variable antes de una segunda es el mismo que produce al presentar a la

segunda antes que la primera. Es decir, los diseños contrabalanceados sostienen la suposición de que no existe una transferencia diferencial (asimétrica) entre las condiciones. De no justificarse esta suposición habría interacciones entre las variables que provocaría dificultades en el análisis estadístico".⁴²

b) Diseños factoriales: se utilizan cuando pretendemos constatar el efecto producido por varias variables simultáneamente, conociendo lo que se debe a cada una y lo que es producto de la interacción entre ellas. Basados en el análisis de varianza, presentan diversas modalidades (diseños factoriales simples y complejos) exigiendo un estudio más amplio que rebasa los límites de este trabajo.

4. *Restricción de las condiciones en la experimentación.* Es el modo de eliminar variables extrañas. Ello ha llevado a la investigación en el laboratorio. Pueden, así, utilizarse cabinas insonorizadas, pantallas que se interpongan entre el experimentador y el sujeto, etc... Sin embargo, en el terreno educativo, las investigaciones en el laboratorio han sido muy criticadas ya que atentan, a veces, a la propia validez de la experimentación.

Aun así, y teniendo en cuenta sus limitaciones, este tipo de investigaciones puede ser utilizado.⁴³

El experimentador como variable extraña

Al hablar de las variables situacionales señalábamos la necesidad de controlar aspectos tales como las normas dadas, presentación de las pruebas, etc... Si a los diversos grupos atienden distintos experimentadores, éstos pueden jugar un papel a la hora de predecir las influencias que han podido recibir los sujetos.

Las fuentes de variabilidad pueden proceder:

- de la desigual impartición de las normas y tratamientos de las condiciones experimentales
- del nivel de "expectativa" y orientación de la misma en orden a la investigación por parte del que la realiza
- de su posible capacidad de conectar (positiva o negativamente) con el grupo de sujetos con los que trabaja

Se han intentado subsanar utilizando equipos mecánicos, midiendo el nivel de expectativa o entrenando a los mismos maestros para que sean ellos quienes lleven a cabo el paso de prueba. En cada experimento habrá que estudiarse cuidadosamente este factor.

42. Mc Guigan: *Obra cit.*, p. 165.

43. Ver especialmente sobre esto el c. 12 de Travers: *Obra cit.*, pp. 331 ss.

CONTROL DE LA VARIABLE INDEPENDIENTE

De dos maneras controlamos esta variable: *manipulándola* o *seleccionando* sus valores. Por ejemplo: yo puedo graduar la intensidad de la luz o la cantidad de práctica a realizar en ejercicios de matemáticas; pero no puedo manipular la inteligencia (es decir variar el C.I. de un sujeto a mi antojo o su sexo). "La manipulación intencional ocurre cuando el investigador determina los valores de la variable independiente, "crea" dichos valores por sí mismo y determina qué grupo de sujetos recibirá qué valor. La selección se realiza cuando el investigador escoge sujetos que ya poseen los valores deseados de la variable independiente".⁴⁴ Desde luego, el único enfoque para determinar la relación causa-efecto es mediante el ejercicio de control a través de la manipulación intencional.

Como uno de los principales objetivos del investigador es conseguir que los cambios producidos en la variable dependiente se deban a diversos tratamientos experimentales (variable independiente), se ve claro que uno de los esfuerzos principales tenderá a intentar aumentar al máximo los efectos de esta última, de modo que la varianza primaria de la variable dependiente sea significativamente mayor que la secundaria (influencia de las variables extrañas).

Los procedimientos para facilitar el aumento de la varianza primaria ya han sido en parte expuestos al tratar del control de las variables extrañas. No cabe duda de que, en la medida que éste se lleve a cabo, existirán más probabilidades de una incontaminación de las variables y, por ello, de la aparición clara de su radio de influencia.

Otros que pueden utilizarse es el escoger los valores extremos de la variable independiente. O bien, cuando la relación entre la variable dependiente y la independiente no es lineal sino curvilínea conviene escoger los valores medios. Cuando no se tiene la seguridad de qué tipo de valores debe escogerse para una mayor eficacia en la investigación, conviene elegir una muestra amplia de los mismos, con lo que se logrará una información más adecuada sobre la verdadera relación entre las variables.

CONTROL DE LA VARIABLE DEPENDIENTE

Puede llevarse a cabo a partir de una única medida, mediante dos medidas, antes y después de que la variable independiente haya ejercido su influencia (pretest-postest), utilizando una medida posterior para observar la permanencia del efecto de la variable independiente (prueba diferida) o llevando a cabo sistemáticamente una secuencia de mediciones (investigaciones

44. Mc Guigan: *Obra cit.*, p. 184.

de desarrollo). Cada una de ellas entraña ciertos riesgos que es preciso controlar (efecto del aprendizaje de la prueba, regresión estadística, acostumbramiento a los tests, etc.). En cuanto a la medición de esta variable puede llevarse a cabo controlando:

- la velocidad de respuesta
- la exactitud de respuesta
- latencia de respuesta
-

Y otras medidas combinadas que permiten ofrecer un amplio margen a la creatividad del investigador.

CONTROL DE LA VARIANZA DEBIDA AL ERROR ⁴⁵

La "varianza debida al error" (error variante), es la inconsistente fluctuación de las medidas de la variable dependiente. Esta variación es fruto de una multitud de factores imponderables que hacen variar una puntuación determinada en un sentido o en otro. El investigador deberá reducir al máximo este tipo de error, si pretende obtener una serie de resultados precisos y significativos.

Dos son los principales componentes de la varianza del error. Por un lado tenemos las diferencias individuales, tanto en el modo de reaccionar como de recibir la variable independiente. Los individuos estamos afectados en cada momento por una serie de variables, tanto subjetivas como ambientales, cuyo control es absolutamente imposible.

Otra fuente posible de error, procede de los métodos utilizados para obtener, cuantificar y analizar los datos. Todo ello constituye los errores de medida. Para eliminar, al máximo, el riesgo de que las medidas varíen al azar, deberemos minimizar la varianza del error.

La minimización de la varianza del error, tiene dos objetivos principales: a) reducir los errores de medida mediante condiciones totalmente controladas (con lo que disminuimos la posibilidad de que factores incontrolados operen al azar); y b) aumentar la "exactitud" de las medidas (cuanto más ajustadas y exactas sean las medidas, mejor podremos obtener la varianza sistemática, y reducir así, la varianza debida al error).

Por otra parte, si minimizamos al máximo la "varianza del error" concedemos a la varianza primaria una mayor posibilidad de mostrar su "significancia" (si realmente lo es). Puede darse el caso de que entre dos condiciones existe una verdadera diferencia significativa, pero si hemos utilizado imprecisos sistemas de medidas, o no hemos utilizado un adecuado control, quizá se produzca una gran varianza debida al error, quedando de esta manera

45. Arnau, J.: *Obra cit.*, p. 93.

minimizada la varianza primaria. Recordemos, de nuevo, la fórmula de la variación total:

$$V_t = V_g + V_e$$

La varianza total de las medidas, es igual a la suma de la varianza primaria (entre los grupos) y la varianza debida al error (errores al azar). Si se aumenta la "varianza del error", disminuye el valor de la varianza sistemática, con lo que las posibilidades de que "dicha varianza sea significativa" se reduce totalmente.

Consideremos, por ejemplo, las siguientes pruebas de significación:

$$t = \frac{\text{estadístico}}{\text{error estándar del estadístico}}$$
$$F = \frac{V_g}{V_e}$$

Para que podamos "probar una hipótesis", basados en que los diversos procedimientos hacen variar significativamente los resultados, deberemos conseguir, no sólo una exacta medida de la variación debida a los "procedimientos experimentales" (numerador), sino, a su vez, un muy ajustado y preciso cálculo de los errores al azar (denominador). La significación depende, por tanto, de la magnitud del numerador (y ésta se consigue mediante utilización de unos adecuados valores de la variable independiente), como de la reducción del numerador (mediante la utilización de una técnica que nos ofrezca el máximo control y precisión de las medidas) .

BIBLIOGRAFÍA

- ARY, D. y otros: *Introduction to Research in Education*. Rinehart, Nueva York, 1972.
- CAMPBELL, D. T.: *Experimental and quasi-experimental design for research*. Rand M. Nally, Chicago, 1963.
- EDWARDS, A.: *Experimental design in Psychological Research*. Rinehart, Nueva York, 1950.
- FOX, D.: *The Research process in Education*. Rinehart, Nueva York, 1965.
- KERLINGER, F. N.: *Foundation of Behavioral Research*. Rinehart, Nueva York, 1965.
- LEWIS, P. G.: *Experimental design in Education*. University of London Press, Londres, 1968.
- MC CUIGAN, S. J.: *Psicología experimental. Enfoque metodológico*. Trillas, México, 1971.
- MC NEMAR, G.: *Psychological Statistics*. John Wiley and Son, Nueva York, 1963.
- MOULLY, G. D.: *The Science of Educational Research*. Vand Nostrand Reinhold Company, Nueva York, 1970.
- TORANZOS, F.: *Teoría estadística y aplicaciones*. Kapelusz, Buenos Aires, 1962.
- TRAVERS, R. M.: *Introducción a la investigación educacional*, Paidós, Buenos Aires, 1971.

EL DISEÑO EXPERIMENTAL: SU APLICACIÓN A LAS CIENCIAS DE LA CONDUCTA

por JAIME ARNAU

1. INTRODUCCIÓN

1.1. Diseños simples y factoriales

2. EXPERIMENTOS BIVALENTES, MULTIVALENTES Y FACTORIALES

2.1. Experimentos bivalentes

2.1.1. Diseño "antes y después"

2.1.1.a) Esquema del diseño antes y después

2.1.1.b) Ventajas e inconvenientes

2.1.2. Diseño de "dos grupos al azar"

2.1.2.a) Esquema del diseño de dos grupos al azar

2.1.2.b) Ventajas e inconvenientes

2.1.3. Diseños de "grupos: dos grupos apareados"

2.1.3.a) Esquema del diseño de dos grupos apareados

2.1.3.b) Apareo mediante "variable de correlación"

1.1.3.c) Apareo por "pares"

2.1.3.d) Apareo mediante el "control acoplado"

2.1.3.e) Apareo según "un rendimiento previo"

2.1.3.f) Ventajas e inconvenientes de los diseños de dos grupos apareados

2.1.4. Diseño "Solomón de cuatro grupos"

2.1.4.a) Esquema del diseño "Solomón de cuatro grupos"

2.1.4.b) Ventajas e inconvenientes

2.1.5. Diseño "intra-sujeto con tratamientos contrabalanceados"

2.1.5.a) Esquema del diseño intra-sujeto

2.1.5.b) Ventajas e inconvenientes

2.2. Experimentos multivalentes

2.2.1. Diseño de grupos completamente aleatorizados

2.2.1.a) Esquema del diseño

2.2.1.b) Ventajas e inconvenientes

2.2.2. Diseño de bloques de grupos al azar

2.2.2.a) Esquema del diseño

- 2.2.3. El doble bloqueo
 - 2.2.3.a) Ventajas e inconvenientes del sistema de bloques
- 2.2.4. Diseño de "medidas-repetidas" o "intra-sujeto"
 - 2.2.4.a) Esquema del diseño
 - 2.2.4.b) Ventajas e inconvenientes

2.3. Experimentos factoriales

- 2.3.1. Efectos factoriales (principales y de interacción)
- 2.3.2. Diseño experimental factorial 2x2
- 2.3.3. Principio de la confusión

3. RESUMEN

4. BIBLIOGRAFÍA

	1
1.1. Diseños simples y factoriales	1.1
2. Experimentos factoriales multivariantes y factoriales	2
2.1. Experimentos bivalentes	2.1
2.1.1. Diseño "antes y después"	2.1.1
2.1.1.a) Esquema del diseño antes y después	2.1.1.a
2.1.1.b) Ventajas e inconvenientes	2.1.1.b
2.1.2. Diseño de "dos grupos al azar"	2.1.2
2.1.2.a) Esquema del diseño de los grupos al azar	2.1.2.a
2.1.2.b) Ventajas e inconvenientes	2.1.2.b
2.1.3. Diseños de "grupos: los grupos aprenden"	2.1.3
2.1.3.a) Esquema del diseño de los grupos aprenden	2.1.3.a
2.1.3.b) Ventajas e inconvenientes	2.1.3.b
2.1.3.c) Aprendizaje por "pairs"	2.1.3.c
2.1.3.d) Aprendizaje mediante el "control análogo"	2.1.3.d
2.1.3.e) Aprendizaje según un "medimiento previo"	2.1.3.e
2.1.3.f) Ventajas e inconvenientes de los diseños de los grupos aprenden	2.1.3.f
2.1.4. Diseños "solución de cuatro grupos"	2.1.4
2.1.4.a) Esquema del diseño "solución de cuatro grupos"	2.1.4.a
2.1.4.b) Ventajas e inconvenientes	2.1.4.b
2.1.5. Diseños "intra-sujeto con tratamientos controlados"	2.1.5
2.1.5.a) Esquema del diseño intra-sujeto	2.1.5.a
2.1.5.b) Ventajas e inconvenientes	2.1.5.b
2.2. Experimentos multivariantes	2.2
2.2.1. Diseño de grupos completamente aleatorizados	2.2.1
2.2.1.a) Esquema del diseño	2.2.1.a
2.2.1.b) Ventajas e inconvenientes	2.2.1.b
2.2.2. Diseño de bloques de grupos al azar	2.2.2
2.2.2.a) Esquema del diseño	2.2.2.a

1. INTRODUCCIÓN

No existe en terminología científica un término que se presente tan ambiguo e impreciso como el de "diseño experimental". En efecto, el concepto de "diseño experimental" lo encontramos, frecuentemente, definido de muy diversas maneras. En algunas definiciones el "diseño" es considerado como la "especial disposición que reciben los tratamientos experimentales" dentro de un trabajo de investigación. En otras es concebido, más bien, como una concreta forma de control no sólo de las variables en las que estamos directamente interesados, sino también, del resto de variables que, sistemática o no sistemáticamente, pueden llegar a afectar los resultados. También por diseño se suele entender un "conjunto de prescripciones" que permiten llevar a cabo comparaciones entre los tratamientos y sus posibles combinaciones. De esta forma podríamos ir enumerando las diversas concepciones sobre el diseño, y al mismo tiempo, podríamos comprobar que si bien cada una de estas diversas interpretaciones tienden a destacar una de sus funciones básicas, todas ellas coinciden en la idea de que el diseño es el "mejor procedimiento para llegar a la obtención de la información necesaria para la adecuada solución de los problemas científicos".

Según la teoría clásica del diseño experimental, éste es concebido como aquel procedimiento o técnica que nos señala la forma de asignar los tratamientos a las diversas unidades o grupos experimentales. Según este criterio, el diseño se convierte en un conjunto de reglas o prescripciones que regulan la forma de aplicación de las condiciones experimentales. No vamos a negar aquí la validez de este enfoque tan ampliamente desarrollado por Fisher. Sólo ampliaremos este punto de vista afirmando que el diseño no sólo implica una serie de reglas sobre la aplicación de los tratamientos, sino que además constituye el verdadero plan o estrategia a seguir por el científico. Vamos a considerar, pues, al diseño como un "auténtico esquema de acción", el cual supone, por parte del experimentador, todo un laborioso proceso de toma de decisiones. Es decir, para hacerse con la información necesaria para su hipótesis, deberá plantearse y resolver una serie de cuestiones referentes a:

- 1) la identificación de las variables que, de una manera u otra, van a intervenir en su experimento;
- 2) el tipo de medidas que va a utilizar para el registro de sus observaciones;

- 3) la cuantificación y análisis de sus datos para poderles aplicar las correspondientes pruebas estadísticas de significación;
- 4) la posibilidad de generalización de los resultados.

Todo esto, como es lógico, está supeditado a uno de los objetivos primordiales del diseño. Es decir, verificar hasta qué punto son sostenibles las relaciones que suponemos existen entre las diversas variables experimentales.

Evidentemente la estrategia básica que caracteriza el diseño depende, en última instancia, de la propia naturaleza del problema experimental. Con ello queremos dar a entender que si bien existen unos modelos de diseños "estandarizados", el concepto de diseño tiene un significado mucho más amplio en el sentido de que cada "experimento requiere un tipo de diseño propio". Sin embargo, puesto que la metodología científica ha elaborado una serie de "modelos de diseños básicos" cuya eficacia ha sido demostrada en una gran cantidad de trabajos experimentales, nos vamos a referir a estos modelos como pautas básicas en la presentación del diseño experimental. Hemos de señalar, a su vez, que junto a su estructura lógica, en todo diseño subyace una "estructura matemática", cuyos presupuestos acerca de la naturaleza de las variables y la relación entre las mismas han de admitirse si se pretende una adecuada manipulación matemática del mismo. Estos presupuestos implícitos en los modelos matemáticos del diseño hacen referencia, como hemos indicado, a la naturaleza y a la forma cómo se distribuyen las variables, y también, a cómo se relacionan. Estos dos conceptos son de suma importancia para su operatividad.

Por último, no podemos olvidar que si bien el diseño debe constituir un adecuado "esquema de acción", pueda ser considerado, también, como una buena técnica de control del "error experimental". Desde este punto de vista, sin tener en cuenta una serie de consideraciones de coste y esfuerzo, podemos encontrar entre los diversos diseños que presentaremos notables diferencias en cuanto a la reducción de los errores experimentales. De ahí que la mayor o menor precisión que podamos lograr en la valoración de nuestros resultados depende, en gran parte, de la correcta elección del diseño.

1.1. *Diseños simples y diseños factoriales*

Desde que Fisher (1925, 1935) estableció las bases para el desarrollo de los diseños factoriales, los diseños se han dividido en simples (de factor único) y factoriales (de dos o más factores).

La idea básica de esta clasificación estriba en si nuestro interés se orienta más bien en conocer hasta qué punto los diferentes niveles de una variable independiente van a alterar nuestros resultados experimentales, o bien, si estamos interesados en conocer la influencia simultánea de dos o más factores. En el primer caso se trataría de valorar el efecto de una sola variable independiente. Esto implicaría un control eficaz de las restantes posibles variables independientes capaces de afectar nuestros resultados. En el caso de diseños

factoriales, se trataría de conocer hasta qué punto la sistemática variación de un factor depende o no de la forma como varían los restantes; es decir, su interacción.

Realmente existe una razón muy poderosa que nos lleva a utilizar, siempre que ello sea posible, los diseños factoriales; aunque, desgraciadamente, hasta el momento presente, es escasa su utilización en las ciencias de la conducta.

Si realizamos investigaciones que impliquen diseños de factor único, deberemos estudiar la acción de las variables de forma aislada e independientemente unas de otras. Es decir, tendremos que ir probando, una tras otra, la acción por separado de cada variable, sin que ello nos permita llegar a conclusiones claras de cómo, en una situación dada, actuarían todas ellas si se presentaran simultáneamente. Mediante el empleo sistemático de diseños de factor único no podríamos estar seguros de si la acción de todos los factores se podría considerar como la suma de sus efectos probados independientemente. Suele ocurrir, no obstante y más aún en las ciencias que tratan de la conducta, que la relación existente entre las variables no es lineal sino multiplicativa. De ahí que los diseños de factor único posean muy poca capacidad explicativa. Un ejemplo nos puede ayudar a comprender este concepto. Supongamos que estamos empeñados en estudiar el rendimiento escolar de un grupo de muchachos, en función de factores tales como la motivación, la aptitud intelectual y la personalidad. Para ello podríamos emplear tres diseños de factor único, tomando en cada uno de éstos una de las tres variables. En efecto, posiblemente llegaríamos a unas conclusiones en las que a más motivación, a más capacidad, y a más fuerte personalidad, los sujetos rinden más. A partir de estos resultados cabría pensar que la acción de estos tres componentes actúan, sobre el rendimiento escolar, aditivamente. Es decir, rinde más el que está más motivado, el que además tiene mayor capacidad intelectual y a su vez, tiene una personalidad mucho más estructurada. Si consideramos el caso que haber utilizado un diseño factorial para el estudio de este problema, quizás hubiésemos obtenido resultados totalmente diferentes. Es decir, podríamos haber comprobado, si para los sujetos con una personalidad más acusada el aumento del nivel motivacional repercute positivamente o no sobre el rendimiento escolar. También podríamos analizar, si para los sujetos mejor dotados intelectualmente, la motivación y la personalidad les afecta en un sentido positivo. Todo esto nos hace pensar en la mayor eficacia de los diseños factoriales, sobre todo para aquellas situaciones en las que la variable dependiente es efecto de un conjunto de factores interdependientes.

2. EXPERIMENTOS BIVALENTES, MULTIVALENTES Y FACTORIALES

2.1. *Experimentos bivalentes* (Diseños experimentales clásicos)

La razón de esta clasificación de los experimentos se basa en que ésta refleja más fielmente el desarrollo histórico de la aplicación de los diseños

en el campo de la conducta. En efecto, podemos comprobar que la mayoría de los diseños que se emplearon en los inicios de nuestra ciencia fueron en experimentos de tipo bivalente. En estos experimentos, como sabemos, se combina un grupo experimental con otro grupo control, que sirve de criterio de contrastación. Estos diseños, en los que se parte de dos condiciones, tuvieron su cometido no sólo, como señala Andrews (1948), como método de aproximación a los problemas científicos, sino incluso como procedimiento para la identificación de variables. Estos diseños han servido para potenciar nuevas y más sistemáticas investigaciones. Son diseños de tanteo en áreas científicas desconocidas y han demostrado su utilidad, sobre todo, en las fases de constitución de una ciencia.

Por otra parte, hemos de señalar, en relación a los posibles resultados conseguidos mediante la utilización de tales diseños, que las posibilidades de generalización de los mismos son limitadas, y las conclusiones podrán ser aplicadas teniendo en cuenta una serie de limitaciones inherentes.

Vamos a considerar unos casos hipotéticos que nos pueden servir de ilustración para comprender los posibles peligros que encierra este tipo de experimentos, sobre todo en el momento de interpretar los resultados. Partamos del supuesto que estamos estudiando el efecto del "impulso del hambre" (manipulación mediante las horas de privación) sobre la "rapidez de aprendizaje discriminativo". Para ello hemos elegido un valor arbitrario de la variable "impulso", que podría ser de "12 horas de privación", y comparamos su efecto con el de "0 horas de privación" (grupo control en este caso). Los resultados podrían ser representados mediante el gráfico 1.

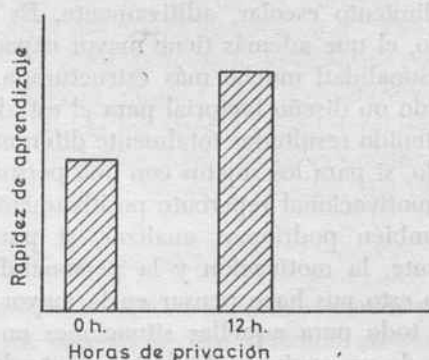


GRÁFICO 1

Según se desprende de este gráfico, es posible que el investigador, si ha comprobado que la diferencia es significativa, concluya que "a mayor privación se dará una mayor rapidez de aprendizaje discriminativo". En principio, su interpretación se basa en datos empíricamente válidos y estadísticamente significativos.

Supongamos, a continuación, que otro investigador, realice un experimento similar, pero que elija, en lugar del valor "12 horas de privación", el valor "32 horas", y haga la comparación con "0 horas de privación". Sus resultados podrían ser representados en la siguiente forma (gráfico 2):

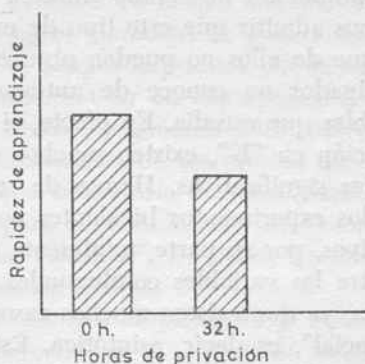


GRÁFICO 2

Comprobada, también, que la diferencia entre ambos resultados es significativa, puede concluir que "a mayor cantidad de horas de privación se opera una disminución en la rapidez del aprendizaje".

Del análisis de estos dos experimentos, correctamente planificados, comprobamos que los dos científicos han llegado a resultados completamente contradictorios. Si intentamos estudiar las posibles causas de esta evidente contradicción, concluiremos que ha habido, por parte de los dos investigadores una "excesiva sobreestimación de sus resultados", dándoles un valor que en realidad no tenían. Por otra parte, ambos investigadores han partido del supuesto, totalmente falso, que la relación existente entre las dos variables es lineal, sin prever la posibilidad de que podía haber sido curvilínea, según podemos observar en el gráfico 3.

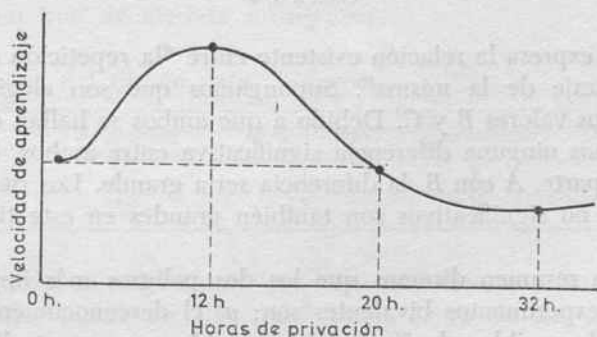


GRÁFICO 3

De la consideración del gráfico anterior, podemos concluir que si un investigador hubiese elegido el valor "20 horas de privación" y lo hubiese comparado con el valor "0 horas", no habría encontrado diferencia significativa alguna entre ambos. Y si, por otra parte, hubiese elegido un valor superior a "32 horas" el rendimiento observado hubiese sido significativamente menor. En suma, debemos admitir que este tipo de experimentos son básicamente exploratorios, y que de ellos no pueden obtenerse conclusiones definitivas, ya que el investigador no conoce de antemano la posible relación existente entre las variables que estudia. En efecto, si las variables presentan entre sí un tipo de relación en "U", existen muchos valores cuyas comparaciones no dan diferencias significativas. Hemos de reconocer, pues, que los resultados positivos de los experimentos bivalentes no son, en absoluto, confirmatorios; ni los negativos, por su parte, totalmente definitivos.

No toda relación entre las variables conductuales se resuelve en función de una línea o una curva, ya que existen muchos casos en que dicha relación es típicamente "exponencial", es decir, asintótica. Ésta suele presentarse en las clásicas curvas de aprendizaje, y en la mayoría de trabajos realizados con estímulos sensoriales. Fijémonos en una función de este tipo, y elijamos, arbitrariamente, tres valores de la misma (gráfico 4).

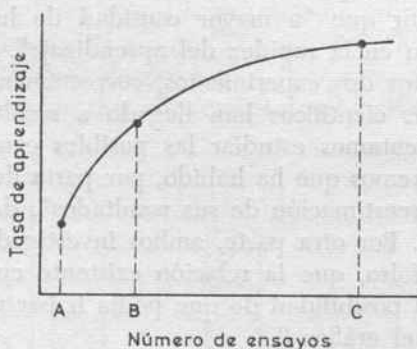


GRÁFICO 4

Esta curva expresa la relación existente entre "la repetición de una tarea" y "el aprendizaje de la misma". Supongamos que son elegidos, para un experimento, los valores B y C. Debido a que ambos se hallan en la asintota, no obtendríamos ninguna diferencia significativa entre ambos. Si comparásemos, por otra parte, A con B, la diferencia sería grande. Los riesgos de seleccionar valores no significativos son también grandes en este tipo de experimentos.

A título de resumen diremos que los dos peligros más importantes que presentan los experimentos bivalentes son: a) el desconocimiento inicial que se tiene sobre la posible relación las dos variables que se estudian. Según los valores que hayamos elegido, podremos llegar a conclusiones totalmente con-

tradictorias sobre unas mismas variables. *b)* Otro de los peligros que corremos se refiere a la generalización de los resultados. Para no caer en el mismo, será suficiente tener en cuenta que toda información que obtengamos en un tipo de experimento bivalente es siempre muy restringida, y, por tanto, tan sólo aplicable a la situación concreta en la que nos hemos movido. Esto nos exige tener mucha prudencia en el momento de generalizar los resultados, no ampliándolos más allá de los límites impuestos por la situación estudiada.

Los diseños que podemos usar en experimentos bivalentes, pertenecen a la categoría de "factor único", y generalmente suelen utilizar el "grupo control" como base de contrastación de los efectos de la variable experimental. Por otra parte, puesto que la mayoría de estos diseños fueron los que primeramente se utilizaron para el estudio de los fenómenos conductuales, los hemos considerado como "clásicos". Es decir, aquellos que han servido para poner las bases de nuestra ciencia.

A efectos prácticos, vamos a emplear, en la sucesiva presentación de los diseños la siguiente nomenclatura. X = variable independiente; O = observaciones realizadas sobre los fenómenos conductuales; Az = asignación aleatoria; Ap = Asignación mediante un sistema de apareo; E = grupo experimental, y C = grupo control.

Vamos a presentar, a continuación, los principales diseños de "factor único" o como hemos señalado anteriormente, "diseños clásicos".

2.1.1. Diseño "antes y después"

En este tipo de diseño se utilizan dos grupos, uno experimental (E) y otro control (C), a los que se les asignan los sujetos mediante un procedimiento al azar (Az). Con ello se pretende conseguir dos grupos que teóricamente sean equivalentes, y esto se consigue mediante la aleatorización. Una vez formados los dos grupos, se mide alguna característica conductual relacionada con la experiencia; medida que nos puede servir como criterio de equivalencia de los grupos. Luego, el grupo experimental es sometido al tratamiento que se estudia, y se pasa a comparar las puntuaciones obtenidas por los dos grupos mediante algún tipo de medida subsiguiente.

2.1.1.a) Esquema del diseño antes y después

Grupo	Asignación	Observ. antes	Tratamiento	Observ. después
E	Ez	O_1	X_1	O_2
C	Az	O_3	X_0	O_4

La base formal de este diseño consiste, mediante la utilización de algún tipo de medida, tomar los datos del rendimiento de los sujetos antes y después que uno de los dos grupos haya sido sometido a un tratamiento experimental (X_1) y el otro no (X_0). Las puntuaciones "antes", como ya hemos

indicado, nos pueden servir para comprobar la equivalencia inicial de los dos grupos con respecto a la variable de medida. Si ésta se demuestra, nos bastará comparar las medidas de las observaciones O_2 y O_4 , para probar nuestra hipótesis.

En el caso de que la equivalente entre los dos grupos no se haya demostrado, será necesario llevar a cabo algún tipo de ajuste de la medida "después" en relación a las iniciales, para poder, así, contrastar la hipótesis. (Para ello se suele emplear, normalmente, alguna forma del análisis de covariancia.)

2.1.1.b) Ventajas e inconvenientes

Gracias a la aleatorización, se puede contar con una técnica de control para que todas las posibles fuentes de error afecten los resultados. El hecho, pues, de que el experimentador puede realizar una prueba de la equivalencia inicial de los dos grupos le permite contar con una base razonable de que ninguna variable haya llegado a interferir con la variable experimental. De este modo, mediante este diseño no sólo se parte del supuesto de que los dos grupos son homogéneos, sino que además se puede verificar la efectividad de dicha técnica de control.

A su vez, cualquier tipo de influencia externa que puedan recibir los grupos se puede estimar mediante la comparación de las puntuaciones "antes y después" de ambos grupos. Al mismo tiempo, dado que las observaciones "antes y después" ocurren en una misma secuencia temporal, poseemos una garantía de que los resultados no han sido alterados como consecuencia de los cambios habidos en la situación.

Una de las limitaciones más importantes en este tipo de diseños, así como para todos aquellos que utilizan la aleatorización como técnica de formación de grupos, es que se requiere una muestra bastante grande de sujetos, para que se puedan controlar, efectivamente, todas las fuentes extrañas de error.

Por otra parte, el hecho de tomar medidas iniciales a los sujetos puede predisponer al grupo experimental para el posible tratamiento que le podamos aplicar.

2.1.2. Diseño de "dos grupos al azar"

Como en el diseño anterior, el investigador asigna los sujetos a los grupos experimental y control, mediante un procedimiento al azar. Esta técnica, como hemos visto, nos garantiza la equivalencia inicial de los dos grupos con respecto a todas las posibles fuentes de error.

2.1.2.a) Equema del diseño de dos grupos al azar

Grupo	Asignación	Observ. antes	Tratamiento	Observ. después
<i>E</i>	<i>Az</i>	—	X_1	O_1
<i>C</i>	<i>Az</i>	—	X_0	O_2

2.1.2.b) Ventajas e inconvenientes

Partiendo de la base, gracias al empleo de la aleatorización, que los dos grupos son equivalentes, el tipo de análisis aplicable a este diseño es relativamente sencillo. Es suficiente verificar, con una adecuada prueba estadística, si la diferencia de puntajes que tenemos de los dos grupos es mayor a la que normalmente cabría esperar si sólo hubiese actuado el azar. Por otra parte, puesto que en este diseño sólo se toman medidas de las observaciones "después" de la aplicación del tratamiento, se elimina el posible "error progresivo" debido al efecto de aprendizaje o fatiga, que podría llegar a confundir los resultados.

2.1.3. Diseño de "grupos apareados"

La idea central de este diseño es obtener dos muestras de sujetos que, en relación a alguna actividad o característica conductual, presenten una media y una desviación estándar lo más idénticas posible. Para ello se asignan al azar, a cada uno de los dos grupos, pares de sujetos con características similares en relación a la variable de conducta. Ello nos asegura una homogeneidad inicial entre los dos grupos y, consecuentemente, nos proporciona una base lógica para inferir que cualquier cambio ulterior que se observe después de la introducción de la situación experimental, será debido a ésta.

Existe, por tanto, una clara ventaja de este procedimiento con respecto al del azar, puesto que al conseguir dos grupos más homogéneos se reduce enormemente el "error experimental" debido a las diferencias individuales (una de las fuentes de error más importantes dentro del área de las ciencias conductuales). Con ello, conseguimos una mayor sensibilidad para detectar valores más débiles de la variable independiente.

Hemos de destacar, asimismo, que este tipo de diseños han sido utilizados, preferentemente, en los llamados "experimentos de laboratorio", ya que con ellos se puede lograr un mayor control de las fuentes de error.

2.1.3.a) Esquema del diseño de dos grupos apareados

Grupo	Asignación	Observ. antes	Tratamiento	Observ. después
<i>E</i>	<i>Ap</i>	—	X_1	O_1
<i>C</i>	<i>Ap</i>	—	X_2	O_0

Vamos a presentar, a continuación, las principales técnicas de apareo que podemos utilizar en este tipo de diseño.

2.1.3.b) Apareo mediante "variable de correlación"

Según esta técnica, el apareo se consigue formando dos grupos con características similares en relación a una variable directamente vinculada con la

variable de media (variable dependiente). Si estudiásemos, por ejemplo, el efecto de práctica (masiva y distribuida) sobre un rendimiento de tipo sensorio-motriz, podríamos formar dos grupos equivalentes, con una media y una desviación estándar igual, si previamente pasamos a los sujetos un test de coordinación viso-motor. De esta manera aseguramos que los dos grupos son equivalentes en relación a la tarea objeto de nuestro estudio.

Muchas veces, la equivalencia se puede conseguir en relación a otras variables como la edad, el sexo, educación, etc. Siempre que, en principio, se piense que están directamente relacionadas con la actividad que va a ser medida.

2.1.3.c) Apareo "por pares"

Existe la posibilidad de formar dos grupos homogéneos asignando al azar, a cada uno de ellos, pares de sujetos con características psicológicas iguales. Así, si por ejemplo deseamos formar dos grupos equivalentes en relación a los cocientes intelectuales, podemos seleccionar pares de sujetos con igual C.I., y a continuación, asignarlos al azar, al grupo experimental y control. También cabría la posibilidad de aparearlos en relación a la edad, sexo, rendimiento escolar, etc.

Hemos de tener en cuenta, que para este tipo de apareo, se forman grupos de igual tamaño.

2.1.3.d) Apareo mediante el "control acoplado"

Durante estos últimos años se está utilizando un tipo de apareo que recibe el nombre de "control acoplado". El origen de esta técnica se remonta a una serie de experimentos en los que se comprobó que si los animales, hambrientos o sedientos, recibían una descarga eléctrica antes de llegar a la recompensa, adquirirían muchas más úlceras de estómago que los que no se les aplicaba (grupo control). Esto planteó el problema de si el origen de la úlcera se debía al choque eléctrico por sí mismo, o bien, a la tensión creada en el animal ante la situación conflictiva de tender hacia la comida y evitar la descarga. Para solucionar este problema, era necesario emplear un grupo control que recibiera la misma magnitud y secuencia temporal de la descarga, pero sin pasar por la situación conflictiva.

Brady (1958), nos presenta la forma como empleó la técnica del "grupo control acoplado". Para ello entrenó unos monos a que presionaran una barra con objeto de evitar una descarga eléctrica en los pies, con una frecuencia de 20 segundos. Esta descarga podía ser evitada si el mono oprimía la barra en el intervalo de los 20 segundos. Los monos aprendieron rápidamente esta conducta, y sólo cuando dejaban de empujar la palanca, recibían una descarga eléctrica.

Para conocer si la úlcera se formaba a consecuencia del choque eléctrico o de la situación conflictiva, se acopló un segundo mono al primero, dentro de la misma situación. La única diferencia consistía en que este segundo

mono, acoplado al primero, recibía las mismas descargas con las mismas intensidad y regularidad que su compañero, sólo esto podía hacer terminar o evitar la aparición del choque. Es decir, los dos recibían la misma descarga, con la excepción de que el sujeto experimental tenía el control de la situación. Esto delimitaba claramente los dos campos, ya que un mono estaba tan sólo sometido a la descarga y el otro a la tensión (y consecuentemente a las posibles descargas).

Esta técnica, a pesar de tener unos inconvenientes graves, está siendo utilizada cada vez más, y se está aplicando, para el estudio del condicionamiento parpebral, el refuerzo secundario, etc. (Gormezano, Moore y Deaux, 1962), con notable éxito. A pesar de esto, no ha dejado de ser criticado.

2.1.3.e) Apareo según "un rendimiento previo"

Muchos experimentadores han utilizado y utilizan los primeros ensayos como criterio de apareamiento. A partir de los resultados obtenidos en los mismos se puede llegar a formar dos grupos homogéneos, pudiendo, incluso, eliminarse aquellos sujetos que se apartan enormemente del grupo.

Una alternativa a este procedimiento implicaría que durante este período de entrenamiento, los sujetos alcanzaran un determinado nivel de eficiencia, en cada uno de los ensayos que se hayan determinado como previos. Así, por ejemplo, los sujetos deben completar un trabajo en 30 segundos, con una variación de más-menos 5 segundos. Si el sujeto alcanza este criterio en el primer, segundo y tercer ensayos, es asignado, a continuación, a uno de los dos grupos al azar. En el caso que no alcance el criterio, es rechazado para el experimento.

2.1.3.f) Ventajas e inconvenientes de los diseños de grupos apareados

Una de las principales ventajas que presentan los diseños de grupos apareados es que, con ellos, se consigue de una forma más precisa la homogeneidad de los dos grupos antes de llevar a cabo el experimento. Por otra parte, mientras que para los grupos formados al azar se necesita un gran número de sujetos para lograr la equivalencia, para los diseños de grupos apareados con 5 a 10 sujetos por grupo, se pueden ya obtener conclusiones válidas.

Debido, pues, a la previa correlación existente entre las medidas que van a ser objeto de comparación, logramos una mayor reducción de la variancia debida al error, y consecuentemente, aumenta la posibilidad de detectar los efectos más débiles de la variable experimental.

Uno de los inconvenientes más graves se presenta, cuando nos vemos necesitados de aparear los sujetos en más de una característica. Ya que es muy difícil que se encuentren sujetos con rasgos similares, lo cual disminuye la posibilidad de formar grupos iguales. Por otra parte, cuando debemos generalizar los resultados, debido a la serie de restricciones que hemos impuesto al seleccionar los sujetos, ésta queda ampliamente limitada.

2.1.4. Diseño "Solomón de cuatro grupos" (Solomón, 1949)

Este diseño viene a ser, en pocas palabras, una combinación de dos tipos de diseños analizados anteriormente: el diseño "antes y después" y el de "dos grupos al azar".

Este diseño nos permite probar, de una forma mucho más eficaz, no sólo el efecto del tratamiento sobre la variable dependiente, sino incluso la posible interacción entre las "medidas antes" y el "tratamiento".

2.1.4.a) Esquema del diseño de Solomón

Grupo	Asignación	Observ. antes	Tratamiento	Observ. después
E	Az	O_1	X_1	O_2
C	Az	O_3	X_0	O_4
E	Az	—	X_1	O_5
C	Az	—	X_0	O_6

Con este diseño el efecto del tratamiento " X_1 " se puede probar comparando: O_2 con O_1 ; O_2 con O_4 ; O_5 con O_6 , y O_5 con O_3 . Si todas estas comparaciones concuerdan entre sí, la posibilidad de inferir que la acción del tratamiento ha sido efectiva, crece enormemente. Tenemos, pues, en un mismo diseño, la posibilidad de obtener cuatro pruebas para verificar nuestra hipótesis.

Asimismo, con un adecuado análisis estadístico, podemos verificar la interacción existente entre el "tratamiento" y la "pre-prueba".

Tabla de datos 2x2

		Tratamientos	
		X_0	X_1
Pre-prueba	Pre-prueba	O_4	O_2
	No pre-prueba	O_6	O_5

GRÁFICO 5

2.1.4.b) Ventajas e inconvenientes

Al utilizar, en este diseño como en los anteriores, la aleatorización para la formación de los grupos, se controlan todas las posibles fuentes de error. A su vez, tenemos dos réplicas de un mismo experimento, pudiendo probar al mismo tiempo, si el "pre-test" tiene una posible influencia sobre los resul-

tados, y si esta situación de “pre-prueba”, puede llegar a interferir o modificar la acción del tratamiento. De ahí que podamos concluir que mediante este diseño se controla el posible efecto de “error progresivo” que puede darse como consecuencia de la pre-prueba que hemos sometido a los sujetos.

Al ser cuatro los grupos, es necesario que nos aseguremos de su equivalencia y de que en la formación de los mismos no haya cualquier otro criterio capaz de afectar sistemáticamente los resultados.

2.1.5. Diseño “intra-sujeto con tratamientos contrabalanceados”

Hemos comprobado cómo mediante la formación de grupos apareados, la variabilidad “entre-sujetos” se reducía enormemente, con la consiguiente reducción del error experimental, y un aumento de la sensibilidad del experimento. Si se piensa en la posibilidad de utilizar el sujeto como grupo único, cabría esperar que se habría conseguido eliminar totalmente la variabilidad “entre-sujetos”, principal causa del error. En este caso “el sujeto actuaría como propio control”, y puesto que habríamos eliminado las diferencias individuales, estos diseños reciben el nombre de “intra-sujetos”.

El diseño “intra-sujeto” permite probar la acción de dos o más tratamientos. Y puesto que todos son aplicados a un mismo sujeto, se suele utilizar la técnica del contrabalanceo como procedimiento para neutralizar el posible efecto acumulativo de los tratamientos.

2.1.5.a) Esquema del diseño intra-sujeto

Tratamientos	X_1	X_2	X_2	X_1
Observaciones	O_1	O_2	O_3	O_4

La disposición de los dos tratamientos (1, 2, 2, 1) obedece a la técnica del contrabalanceo. Es decir, si el tratamiento “2” va detrás del tratamiento “1”, se invierte a continuación tal disposición. Así, los posibles efectos facilitadores o distorsionadores del “tratamiento 1” sobre el “2”, en una primera fase de la prueba, quedan contrarrestados al cambiarse la disposición de los mismos, en la segunda fase del experimento.

Para un análisis de datos, se compararían las observaciones O_2 y O_3 , con las O_1 y O_4 .

2.1.5.b) Ventajas e inconvenientes

Una de las principales ventajas del diseño “intra-sujeto” es la eliminación de la variación debida a las diferencias individuales, lo cual repercute directamente en un mayor control de las variables subjetivas.

Ahora bien, uno de los principales inconvenientes, es la posible acción acumulativa de los factores tales como aprendizaje o fatiga, que pueden llegar a interferir con los tratamientos (error progresivo). Si realmente sospechamos

que existe un efecto facilitador o distorsionador en la aplicación sucesiva de los tratamientos, tendríamos que aplicar la técnica del "análisis de tendencias" propias para diseños "secuenciales". Ahora bien, para un tipo de tratamientos cuyo efecto no es duradero, se puede lograr contrarrestar su acción mediante el contrabalanceo de las condiciones experimentales. Al disponer, pues, el experimento según la pauta, 1, 2, 2, 1, y al calcular la media para ambas observaciones dentro de cada una de estas situaciones, se puede controlar el posible efecto secuencial.

Por otra parte, este diseño implica una menor cantidad de sujetos, ya que sólo se emplea un solo grupo, pues los sujetos actúan dentro de él como su propio control. Debido al efecto, relativamente corto, de los determinantes experimentales, los datos de todas las observaciones se obtienen en una misma sesión de recogida, lo cual evita la pérdida de los mismos por ausencia de los sujetos.

Uno de los presupuestos básicos de este diseño es que las variables secundarias extrañas relacionadas con el tiempo presenten una relación lineal. Si el efecto de tales variables sobre la conducta no refleja una relación lineal, cualquier modificación o cambio de una prueba a otra puede interferir con los efectos de las variables experimentales. En este caso la prueba de este diseño no es adecuada para la situación experimental descrita.

2.2. Experimentos multivalentes

Las técnicas de análisis, perfeccionadas sobre todo por Fisher (1923), han permitido el empleo de diseños de experimentos que aunque impliquen un solo factor (una sola variable independiente) se pueden utilizar dos o más valores de la misma, con objeto de obtener una información más precisa sobre la posible relación existente entre dos variables. Ésta es la razón por la que a tales experimentos se les conoce también con el nombre de "experimentos funcionales".

El gráfico 6 representa un experimento funcional, reflejando la relación entre las variables "resistencia a la extinción de una respuesta aprendida" y "cantidad de refuerzos recibidos durante la adquisición". Como podemos comprobar en este experimento se han utilizado seis valores de la variable independiente: 0, 10, 20, 40, 80 y 160 refuerzos.

Estos valores son totalmente arbitrarios y se podía haber realizado otro experimento tomando como valores, los siguientes: 0, 15, 30, 60, 120 y 240. El científico puede elegir éstos u otros valores, ya que él sólo se halla interesado en la posible línea que unirá estos valores, reflejando la relación existente entre ambas variables. Esta técnica permite, a su vez, interpolar valores que no han sido probados experimentalmente así como predecir, también, para un determinado valor de la variable independiente cuál será la tasa de respuesta dada por el sujeto.

Estos experimentos tienen una gran importancia para la Psicología, no

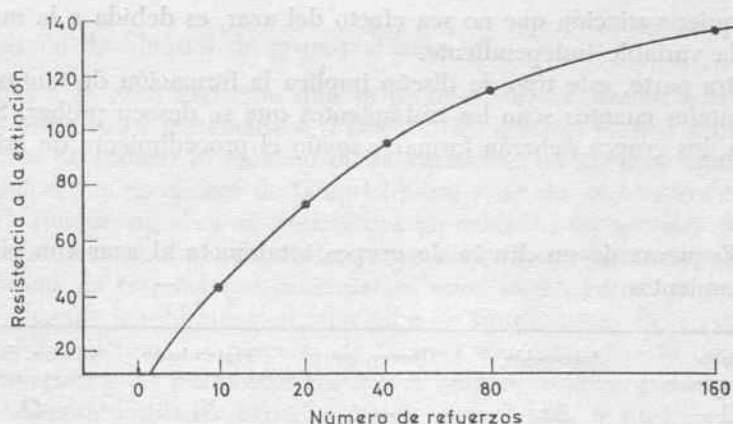


GRÁFICO 6

sólo porque nos permite establecer, con cierta precisión, la posible relación existente entre dos variables, ofreciéndonos una base razonable para predecir el posible efecto de un valor no probado, sino porque a su vez nos permite realizar y probar relaciones entre variables de tipo cuantitativa.

Pertencen a esta categoría de experimentos los siguientes diseños: Diseños de grupos completamente aleatorizados, diseño de bloques de grupos al azar, diseños cuadrados, y los diseños de medidas repetidas (o diseño intra-sujeto). Hemos de tener en cuenta que los diseños aquí señalados no son exclusivos para el tipo de experimentación "multivalente", ya que también son válidos para experimentos "factoriales".

La idea básica que nos guía al dividir los diseños en las categorías antes expuestas, estriba en que cada uno de ellos aplica una técnica diferente para conseguir la reducción del "error experimental".

A continuación vamos a presentar, brevemente, cada uno de estos diseños que, como ya hemos señalado, tanto pueden aplicarse a los experimentos funcionales como a los factoriales.

2.2.1. Diseño de grupos completamente aleatorizados

Muchas veces se ha hablado del diseño en relación a la técnica que emplea para conseguir grupos o muestras de individuos lo más homogéneos posible. Sólo sobre la base de la equivalencia de los grupos se pueden llevar a cabo las pruebas pertinentes sobre las hipótesis experimentales. Por esa razón, los experimentadores han utilizado diversos procedimientos, uno de los cuales ya lo hemos mencionado anteriormente: la aleatorización.

Si la aleatorización ha sido llevada a cabo de una forma efectiva, nos asegura la total neutralización de cualquier tipo de variable que pueda llegar a afectar los resultados. Con ello poseemos una base razonable para concluir

que cualquier variación que no sea efecto del azar, es debida a la manipulación de la variable independiente.

Por otra parte, este tipo de diseño implica la formación de tantos grupos experimentales cuantos sean los tratamientos que se deseen probar. Siempre, claro está, los grupos deberán formarse según el procedimiento de asignación aleatoria.

2.2.1.a) Esquema de un diseño de grupos totalmente al azar con cinco tratamientos

Grupo	Asignación	Observ. antes	Tratamiento	Observ. después
1	Az	—	X_1	O_1
2	Az	—	X_2	O_2
3	Az	—	X_3	O_3
4	Az	—	X_4	O_4
5	Az	—	X_5	O_5

2.2.1.b) Ventajas e inconvenientes

Mediante la aleatorización se logra equilibrar todas las posibles influencias extrañas tanto en un sentido como en otro. Con ello podemos esperar que los datos que obtengamos de los sujetos se distribuyan según la "ley normal", que como sabemos constituye una de las condiciones básicas para la aplicación de las pruebas de significación estadística. No es necesario que el número de sujetos que formen cada grupo experimental sea exactamente el mismo, siempre que el número sea lo suficientemente grande como para que el azar pueda actuar.

Este diseño se basa, como ya hemos dicho, en la técnica de asignación al azar de los individuos a los grupos experimentales. Y esto es, sin duda, uno de sus inconvenientes más grandes. No por el hecho de la utilización del azar, sino porque la variación interindividual será mucho mayor que la de cualquier otro tipo de diseño. Por otra parte, sabemos que esta variación constituye uno de los componentes más importantes de error. Esto, lógicamente, dificultará la posibilidad de probar la eficacia de los tratamientos, sobre todo si la diferencia existente entre ellos es muy débil. Es decir, a mayor error experimental, menor posibilidad de probar la hipótesis (en el caso de que ésta sea válida).

La aleatorización se puede llevar en su práctica más extrema, si asignamos, también, los tratamientos a cada uno de los grupos según un criterio al azar. Ello aumenta nuestra seguridad de que no hay ninguna razón previa por la que un sujeto pueda recibir un tratamiento u otro; sólo el azar es el único responsable.

2.2.2. Diseño de bloques de grupos al azar

Debido a la gran variación que se produce en los diseños totalmente al azar, los estadísticos matemáticos (Fisher, 1923) crearon nuevos procedimientos capaces de reducir al máximo dicha variación. La primera aplicación de estas técnicas fue en el área de la agricultura, y de ahí, el origen del término "bloque". Bloque significa en agricultura un conjunto de parcelas de terreno de características similares (fertilidad, humedad, climatología, etc.). Así, todas las porciones de terreno que presentaban unos rasgos comunes, se consideraron totalmente equilibradas en relación a los tratamientos. En efecto, dentro de un mismo bloque, las parcelas de terreno son entre sí más homogéneas, que en relación a las pertenecientes a otro bloque. Mediante este sistema, se puede conseguir unidades experimentales equivalentes, y de escasa variabilidad. Estos bloques, como es lógico, debido a la homogeneidad interna son más sensibles a las más pequeñas diferencias entre los tratamientos.

El principio del bloque, lógicamente, puede aplicarse a las ciencias biológicas, y, sobre todo, a las de la conducta. De esta manera, si logramos formar bloques de sujetos que presenten características psicológicas similares habremos reducido la variabilidad "entre-sujetos", lo cual nos permitirá formar grupos de sujetos homogéneos. Naturalmente, dentro de cada bloque los grupos deberán ser formados según un procedimiento al azar, así como la aplicación a los mismos de los diversos tratamientos experimentales.

Supongamos que nos interesa estudiar la eficacia de tres métodos de enseñanza, A, B y C y para ello formamos dos bloques de sujetos. A un primer bloque colocamos aquellos sujetos que poseen un coeficiente intelectual de 110, y al otro, sujetos con un C.I., de 90. Dentro de cada bloque la homogeneidad de los sujetos con respecto a la inteligencia es completa. Esto nos va a permitir estudiar con más precisión la eficacia de los tratamientos.

2.2.2.a) Esquema del diseño de bloques de grupos al azar

Grupo	Asignación	Observ. antes	Tratamiento	Observ. después
I	1	—	X_a	O_a
I	2	—	X_b	O_b
I	3	—	X_c	O_c
II	1	—	X_a	O_a
II	2	—	X_b	O_b
II	3	—	X_c	O_c

Este diseño tiene mucha aplicabilidad en el campo de la conducta debido a que las medidas conductuales fluctúan enormemente de un sujeto a otro, por la existencia de las diferencias individuales. Con este procedimiento

conseguimos diseños con mayor precisión que los que utilizan solamente la técnica del azar, y damos una mayor posibilidad a la hipótesis de ser demostrada.

2.2.3. El doble bloqueo

Mediante la técnica del bloqueo, como acabamos de ver, reducimos enormemente el error experimental al agrupar sujetos semejantes en relación a una característica directamente vinculada con los datos de respuesta que nos interesan. Se puede pensar en la utilización simultánea de dos sistemas de bloques. En principio se podrían utilizar dos sistemas de dimensiones diferentes, pero debido a que esto supondría una serie de análisis estadísticos complejos, se ha optado por sistemas de bloqueo que se correspondan entre sí, y asimismo que guarden simetría con la cantidad de tratamientos; de esta manera, se forman matrices cuadradas (dándose una misma cantidad de bloques en un sistema y en otro). Esta clase de diseños ha recibido el nombre genérico de "Cuadrado latino", recogiénolo de un antiguo pasatiempo de carácter matemático (gráfico 7).

Un diseño "Cuadrado latino" está formado por dos sistemas de bloques de n categorías o grupos, cada uno. De tal manera que cada bloque de un sistema contiene "una unidad experimental" perteneciente a cada uno de los bloques del otro. Estos dos sistemas son conocidos, corrientemente, con el nombre de "filas" y "columnas". Así, cada "unidad" o "grupo experimental" se encuentra una sola vez en cada fila y una sola vez en cada columna. Debido a esta especial disposición del diseño sólo se puede utilizar, como ya hemos indicado, una cantidad de tratamientos que sea igual al número de categorías por sistema.

Siguiendo con el ejemplo anterior, podemos suponer que hemos decidido establecer un sistema de tres bloques según la característica "aptitud intelectual". En efecto, se pueden formar tres bloques: el primero para aquellos sujetos que posean un C.I. de 85, el segundo para sujetos con 100 de C.I., y el tercero, para los que tengan un C.I. de 115. A continuación, decidimos agrupar los sujetos según una segunda característica psicológica, por ejemplo, la "capacidad de retención"; para ellos seleccionamos tres percentiles: 30, 50 y 70. Con esto disponemos ya de dos sistemas de bloqueo; cada sujeto sólo se encontrará una y sólo una vez en una fila y en una columna.

Hecho esto, a continuación se asignan, al azar, dentro de cada fila y cada columna, los tratamientos experimentales.

El plan para un diseño de Cuadrado latino, para el ejemplo propuesto, quedaría dispuesto de la siguiente forma:

Cada tratamiento (representado por las letras A, B y C) aparece una y sólo una vez en cada fila y en cada columna. Como hemos indicado, la asignación de los tratamientos dentro de cada bloque es totalmente aleatoria. No existe, pues, por parte de los sujetos una previa disposición para recibir

		Diseño Cuadrado latino		
		Cocientes intelectuales		
		85	100	115
Retención	30	A	B	C
	50	C	A	B
	70	B	C	A

GRÁFICO 7

un tratamiento más bien que otro. El azar constituye, de nuevo, el criterio de confrontación.

Hemos presentado un ejemplo, muy frecuente en Psicología, en que la pertenencia de un individuo a un grupo viene dada por la posesión de unas características aptitudinales dadas. Se podría pensar en otro tipo de diseños, en que la asignación de los sujetos a los grupos se pudiera realizar totalmente al azar. Esto ocurriría si las variables que forman el sistema de bloques fueran externas al sujeto y no orgánicas como hemos visto en el ejemplo anterior. Así se podría estudiar la eficacia psico-motriz en función del tiempo de práctica (5, 10 y 15 minutos de práctica), formando los bloques en función del momento en que se ejecuta la prueba (mañana, tarde y noche), y en función del día de la semana (lunes, miércoles y viernes). En este caso, los sujetos se asignarían, dentro de cada bloque, a cada uno de los grupos mediante un procedimiento aleatorizado.

2.2.3.a) Ventajas e inconvenientes del sistema de bloques

Las ventajas que nos ofrece el sistema de bloques, simple o doble, es sin duda, de gran importancia. Se consigue formar grupos de sujetos similares con respecto a una o dos categorías, lo cual supone una enorme reducción de la variancia debida al error, ya que se separa de la variancia del error la variancia debida a los bloques. El error experimental se calcula, solamente, a partir de la variancia conjunta dentro de cada bloque. Así se logra una mayor precisión en el momento de comparar el efecto de los tratamientos.

Los principales inconvenientes se deben a que los grupos han de formarse con una cantidad igual de sujetos, y, sobre todo, en Psicología, es difícil encontrar sujetos con características similares para utilizar estos diseños. No obstante, es conveniente su uso en Psicología, ya que constituye uno de los métodos más seguros de experimentación.

Por otra parte, hemos de tener en cuenta que si bien mediante el sistema de bloques, reducimos enormemente el error experimental, no por ello se logra eliminarlo totalmente, ya que los sujetos por más que presenten carac-

terísticas similares, dentro de la situación experimental es de esperar que no todas reaccionen de una forma igual y que, por tanto, aparezcan diferencias individuales.

2.2.4. Diseños de "medidas-repetidas" o "intra-sujeto"

Como ya hemos indicado, las diferencias individuales constituyen una de las fuentes de error más importante. El objetivo de las técnicas que hemos ido estudiando ha sido separar dicha variación de la variación debida a los tratamientos y de la variación del error, con lo que se aumentaba la sensibilidad del experimento. Si esta fuente de error no puede ser controlada de alguna manera pasa, automáticamente, a formar parte del error experimental.

Por esta razón, uno de los principales motivos en el uso de este tipo de diseños estriba en el hecho de la eliminación de las diferencias individuales. En efecto, el sujeto es sometido a todos los tratamientos que constituyen el experimento, con lo cual se convierte en su propio control. Así los efectos de un tratamiento concreto son valorados al compararlos con el promedio de respuesta dada por el sujeto en todos los tratamientos. Se ha eliminado, por tanto, de dicho promedio el posible error debido a las diferencias. Puesto que utilizamos el mismo sujeto para las diversas condiciones experimentales, se debería realizar un número de observaciones igual al de los tratamientos, y por esta razón, el diseño recibe el nombre de "medidas-repetidas" o "intra-sujeto", puesto que se ha eliminado la variación entre los sujetos.

2.2.4.a) Esquema del diseño "de medidas repetidas"

Sujetos	Tratamientos				Observaciones			
1	X_1	X_2	X_3	X_4	O_1	O_2	O_3	O_4
2	X_1	X_2	X_3	X_4	O_1	O_2	O_3	O_4
3	X_1	X_2	X_3	X_4	O_1	O_2	O_3	O_4
4	X_1	X_2	X_3	X_4	O_1	O_2	O_3	O_4

2.2.4.b) Ventajas e inconvenientes

Este diseño es sumamente útil, siempre y cuando los efectos de los tratamientos tengan una duración limitada y, por consiguiente, éstos puedan ser aplicados en un corto período de tiempo. Generalmente este tipo de diseño se usa para tratamientos secuenciales, como en las situaciones de aprendizaje, así como en el estudio del efecto de diversas drogas, etc., siempre que la acción de sus efectos pueda ser eliminada rápidamente.

Entre los inconvenientes más importantes tenemos "el error progresivo", o efecto pro-activo de la práctica. Esta acción puede operar positiva o negativamente a la práctica repetitiva, como un efecto facilitador como consecuencia del aprendizaje que se opera.

En el caso de que los diferentes tratamientos guarden entre sí una determinada proporción y nos interese conocer el tipo de relación que existe entre la variable independiente y la dependiente (lineal, cuadrática, cúbica, etc.) podremos aplicar a este diseño el "análisis de tendencias" y probar así hasta qué punto los tratamientos se relacionan según una relación lineal o no.

Por último, como ya hemos indicado, el efecto de los tratamientos deberá ser temporal, sobre todo si se trata de drogas, para evitar la posible confusión entre los tratamientos.

2.3. Experimentos factoriales

Los diseños factoriales son aquellas planificaciones experimentales en las que se estudia simultáneamente la acción de dos o más variables independientes. Cada una de estas variables independientes o factores suele tener varios valores, siendo el diseño factorial, en definitiva, todas las posibles combinaciones entre los valores de una variable con los de la otra. De ahí que, consiguientemente, en un diseño factorial (en el caso de que sea completo) se dan tantos grupos experimentales como posibles combinaciones se puedan realizar entre los diversos tratamientos experimentales. Es, por tanto, imprescindible para que tengamos un "diseño factorial" que al menos estén presentes dos variables independientes.

Son obvias las ventajas que nos reporta el uso de los diseños factoriales y que brevemente vamos a comentar. En primer lugar, nos permite un considerable ahorro de coste, tiempo y sujetos. Quizá se comprenda mejor mediante un ejemplo: si queremos probar la eficacia de tres factores, como "rasgo de personalidad", "agudeza visual" y "capacidad intelectual" sobre la percepción taquistoscópica, tendríamos que repetir tres veces el experimento, en el caso que utilizásemos diseños de factor único. En el primero probaríamos la acción del rasgo de personalidad; en el segundo la agudeza visual, y en un tercero, la capacidad intelectual, lo cual supondría realizar tres experimentos paralelos, triplicar el número de sujetos y aumentar, consecuentemente, el coste del mismo. Mediante un diseño factorial reducimos enormemente el factor tiempo, el número de sujetos y el esfuerzo; además, la información que nos ofrece el diseño factorial, tiene el mismo valor para cada uno de estos factores que si se hubiesen realizado tres experimentos por separado. Por otra parte, no sólo este diseño tiene la misma fuerza probatoria que los diseños simples, sino que además, como cada tratamiento consiste en la combinación de los valores de una variable con respecto a los de las restantes, podemos estudiar al mismo tiempo si la acción de los diversos niveles de un factor varía a medida que varía un segundo factor, o bien se mantiene al margen de su variación. Esto nos ofrece una mayor posibilidad de generalización y obtener una información mucho más valiosa que si estuviésemos trabajando con diseños simples. Así, en lugar de mantener neutralizadas una serie de variables sistemáticas importantes, las introducimos en la situación

experimental y podemos estudiar el comportamiento de nuestra variable en función del comportamiento de las restantes. De esta manera podremos concluir si el efecto de un valor varía a través de los diversos valores de las variables restantes.

Por último, hemos de señalar una tercera ventaja a favor de este diseño en relación a los diseños simples, sobre todo cuando se aplica a los fenómenos conductuales. La conducta es un fenómeno muy complejo que constantemente se halla sometido a la acción de una gran cantidad de variables; mediante la utilización de los diseños factoriales contamos con un procedimiento que nos permite estudiar la variación de la conducta en función de la acción conjunta de varios factores, lo cual nos permite estudiar dicho fenómeno desde una perspectiva más real que lo que se conseguiría si sólo se realizaran diseños simples.

Resumiendo afirmaremos que el diseño factorial supone, junto a un positivo ahorro de sujetos y esfuerzo, un aumento del poder informativo de los resultados. Todo esto sitúa a este tipo de diseños en un lugar privilegiado para el estudio de los fenómenos conductuales.

2.3.1. Efectos factoriales (principales y de interacción)

En todo diseño factorial se combinan dos o más variables independientes denominadas, también, factores, a dos o más niveles. Cada una de las posibles combinaciones recibe el nombre de "tratamiento" y se aplica a una unidad o grupo experimental. Así, pues, cada grupo experimental representa una determinada combinación de un valor de un factor con el de otro factor. El efecto de una variable independiente para cada uno de los valores de la otra, recibe el nombre de "efecto simple"; por el contrario, el efecto total de una variable independiente, promediado a través de todos los valores de la otra recibe el nombre de "efecto principal" de la variación en cuestión.

El efecto simple de una variable a través de la variabilidad de los valores de la otra, viene definido con el nombre de "efecto de interacción" o "efecto secundario". Si tenemos, por ejemplo, la variable independiente A , cuyos efectos varían en función del cambio de los valores de la variable B , se trataría en este caso de un "efecto de interacción de $A \times B$ ", lo cual, también, sería válido para el caso contrario.

Los efectos de interacción, en los que sólo intervienen dos variables, reciben el nombre de "primer orden"; cuando intervienen tres variables, de "segundo orden", etc. (de primer orden, $A \times B$; de segundo, $A \times B \times C \dots$).

2.3.2. Diseño experimental factorial 2×2

Con objeto de estudiar cómo se calculan los "efectos principales y secundarios" de un diseño factorial, vamos a analizar un plan factorial simple. En él se combinan dos variables independientes, A y B , con dos niveles cada una (de ahí, el por qué se denomine diseño factorial 2×2 , o también, 2^2 , que

significa dos variables a dos niveles cada una). Las correspondientes letras minúsculas se utilizan para representar los diferentes niveles de cada factor. Empleando la simbología de Yates (1937), para una mayor simplicidad, tendremos que la presencia de la letra minúscula representará el valor más alto de una variable, y su ausencia el valor más bajo (como cuando son comparadas dos dosis de drogas, cantidad de práctica, etc.). Si tenemos, por ejemplo, la combinación *a*, significará un tratamiento formado por el nivel más bajo de *B* y el nivel más alto de *A*; en el caso que se combinen los dos niveles bajos de *A* y *B*, se representará por el símbolo (1). En la figura siguiente tenemos un esquema de un experimento factorial 2 x 2, en donde se encuentran específicas las diferentes combinaciones posibles entre los diversos niveles de *A* y *B*.

Plan de un experimento factorial 2x2

		A	
		1	2
B	1	(1)	a
	2	b	ab

GRÁFICO 8

Cada uno de los "cuadrados o celdas" de la tabla, viene a representar las cuatro posibles combinaciones de tratamientos, a los que van a asignarse una cantidad *n* de sujetos al azar.

Si simbolizamos el valor medio de las observaciones mediante la letra \bar{X} , podremos calcular los efectos principales de *A* y *B*, de la siguiente forma:

$$\text{Efecto principal de } A = 1/2 [(\bar{X}_{ab} - \bar{X}_b) + (\bar{X}_a - \bar{X}_{(1)})]$$

$$\text{Efecto principal de } B = 1/2 [(\bar{X}_b - \bar{X}_{(1)}) + (\bar{X}_{ab} - \bar{X}_a)]$$

También se podría calcular comparando las medias combinadas de los cuadrados (1) y *b*, con las combinaciones de las medias en *a* y *ab*, para el efecto principal de *A*; siguiendo el mismo procedimiento para calcular el efecto principal de *B* (comparando la combinación de las medias (1) y *a*, con las combinaciones de las medias en *b* y *ab*).

$$\text{Cálculo del efecto de interacción } A \times B = 1/2 [(\bar{X}_a - \bar{X}_{(1)}) - (\bar{X}_{ab} - \bar{X}_b)]$$

Para computar, pues, la interacción entre *A* y *B*, se calcula el efecto simple de *A* para un valor dado de *B* (a través de la diferencia entre la media de *a* y (1), y a continuación se calcula el mismo efecto simple de *A*, para el segundo valor de *B* (a través de la diferencia entre las medias de *ab* y *b*).

La diferencia entre las diferencias nos da el valor de la interacción $A \times B$. Se podría calcular, también, dicha interacción calculando los efectos simples de B para los dos valores de A , y calcular a continuación la diferencia.

$$A \times B = 1/2 [(\bar{X}_b - \bar{X}_{(1)}) - (\bar{X}_{ab} - \bar{X}_a)]$$

Tenemos, pues dos expresiones que nos sirven para el cálculo de la interacción, que no es otra cosa que la diferencia de los efectos simples de una variable. Se pueden simplificar algebraicamente cada una de estas dos expresiones de la siguiente forma:

$$A \times B = 1/2 [(\bar{X}_{(1)} + \bar{X}_{ab}) - (\bar{X}_a + \bar{X}_b)]$$

De esta forma podemos comprobar que la interacción de $A \times B$ consiste en la comparación cruzada de dos pares de tratamientos con los otros dos, o sea la comparación de las combinaciones de tratamientos que se hallan en las diagonales del plan experimental.

2.3.3. Principio de la confusión

Suele ocurrir frecuentemente, sobre todo en Psicología, que al definir una variable mediante la selección de dos grupos de sujetos distintos como podría ser el "nivel socioeconómico" y nos interesásemos por el estudio del efecto de dicha variable sobre las "preferencias vocacionales", ésta puede acusar la influencia de otras variables como, por ejemplo, "las posibilidades ocupacionales futuras", "el escaso nivel de estudios", y que puede hacer variar sistemáticamente nuestra variable dependiente (preferencias vocacionales). Cabe esperar que los sujetos de estratos sociales inferiores consideren sus posibilidades de trabajo, para determinadas profesiones, más restringidas y, a su vez, éstas pueden influir más poderosamente en su elección vocacional que el propio nivel socioeconómico. En este caso nos encontraríamos con más de una variable independiente, que en nuestro ejemplo serían "las expectativas subjetivas", capaces de contaminar totalmente la variable experimental.

Tenemos, pues, lo que suele conocerse con el nombre de "variable confundida". En efecto, siempre que a la variación de los niveles o categorías de una variable independiente se halle asociada la variación de otra variable no controlada, nos encontramos ante un "efecto de confusión". De hecho, uno de los objetivos básicos del control consiste, precisamente, en evitar este tipo de efectos.

Ahora bien, en muchos casos en que debemos utilizar diseños factoriales con un número de tratamientos bastante grande (de 12 o 16 combinaciones), puesto que necesitaríamos una gran cantidad de sujetos para la formación de un bloque o una réplica completa del experimento, muchas veces se opta por reducir el tamaño del bloque o réplica, para reducir el error experimen-

tal. En este caso se sacrifican deliberadamente algunas de las interacciones, en orden a conseguir una mayor precisión en la estimación de los efectos principales. Analicemos este principio mediante un ejemplo muy sencillo, y que en la práctica, no se utiliza, pero que presentamos aquí a efectos metodológicos. Tomemos el diseño experimental 2×2 , y dividámoslo en dos bloques de la siguiente forma:

Bloque I	Bloque II
(1)	a
ab	b

Para el cálculo del efecto principal de A, computaríamos la diferencia entre la media en *ab* y (1) para el primer bloque, y la diferencia entre la media en *a* y *b*, para el segundo bloque, y las sumaríamos. Los efectos de A, se hallarían equilibrados con respecto a los efectos de B.

$$\text{Efecto principal de A} = 1/2 [(\bar{X}_{ab} - \bar{X}_{(1)}) + (\bar{X}_a - \bar{X}_b)]$$

y para el cálculo de efecto principal de B, tendríamos:

$$\text{Efecto principal de B} = 1/2 [(\bar{X}_{ab} - \bar{X}_{(1)}) + (\bar{X}_b - \bar{X}_a)]$$

calculemos a continuación el efecto de la interacción *AxB*:

$$\text{Efecto interacción AxB} = 1/2 [(\bar{X}_{ab} - \bar{X}_b) - (\bar{X}_a - \bar{X}_{(1)})]$$

dicha expresión es matemáticamente equivalente a la siguiente:

$$AxB = 1/2 [(\bar{X}_{ab} + \bar{X}_{(1)}) - (\bar{X}_a + \bar{X}_b)]$$

Tenemos, pues, que para el cálculo de la interacción *AxB*, hemos hallado el total de las medias de observaciones del bloque I, menos el total de las medias del bloque II. Con ello podemos comprobar que la comparación de tratamientos requeridos para el cálculo de la interacción coincide con la de los dos bloques. En este caso se dice que la interacción *AxB*, queda confundida con el efecto de los bloques I y II.

Fijémonos que el cálculo de los efectos factoriales para un tipo cualquiera de diseño 2^n (*n* factores con dos valores cada uno) consiste en la comparación de la mitad de las medias de las observaciones con la otra mitad. Es decir, siempre una mitad se presenta con el signo más y la otra con el signo menos. Así, si tenemos un experimento factorial *ABCD* (de dos valores cada uno), la realización de una réplica completa supondría la utilización de 16 condi-

ciones o combinaciones de tratamientos. Supongamos que hemos asignado n sujetos al azar a cada uno de los 16 grupos, y debido a que la prueba es bastante compleja, deberían realizarla a lo largo de varios días, de manera que algunos sujetos deberían ser probados por la mañana, y otros por la tarde.

Si aleatorizamos la variable "mañana y tarde" de manera que cada sujeto realice su tarea en cada uno de los días en que dure la prueba por la mañana o tarde, según un orden al azar, puede ocurrir que el error experimental aumente considerablemente. En este caso podríamos formar dos bloques de sujetos más homogéneos, de manera que los que pertenecen a un bloque sólo realizaran las pruebas por la mañana, y los del otro, por la tarde. Pero en este caso, tendríamos dentro de cada bloque de la mitad de sujetos. Tendríamos, pues, en cada bloque menos sujetos de los que serían necesarios para todo el experimento. Para solucionar este problema se puede proceder de la siguiente forma: si se parte del supuesto que la "variable mañana-tarde" no es muy importante para nuestro experimento, la podemos sacrificar confundiéndola con la interacción de orden superior $AxBxCxD$, debido a la escasa importancia psicológica que tendría para nuestro experimento.

Partiendo de la base que vamos a desprestigiar la interacción $AxBxCxD$, dividiremos el experimento en dos bloques, de manera que cada uno esté formado por 8 condiciones (la mitad). Así, formaríamos un bloque con los grupos que en la comparación de dicha interacción aparezcan con el signo "más", y el otro, con los grupos que van a tener el signo "menos". Para conocer el signo de cada grupo, en una interacción cuádruple, basta aplicar la siguiente regla (Cochran y Cox, p. 188):

$$AxBxCxD = 1/2^{n-1} (a-1)(b-1)(c-1)(d-1)$$

Desarrollando este polinomio y agrupando los términos positivos a un lado y los negativos a otro, tendremos formados los dos bloques de la siguiente forma:

Cálculo de la interacción $A \times B \times C \times D$. Comparación

+	—
<i>abcd</i>	<i>abc</i>
<i>ab</i>	<i>abd</i>
<i>ac</i>	<i>acd</i>
<i>ad</i>	<i>bcd</i>
<i>bc</i>	<i>a</i>
<i>bd</i>	<i>b</i>
<i>cd</i>	<i>c</i>
(1)	<i>d</i>

La comparación de las condiciones para el cálculo del efecto de interacción cuádruple viene esquematizado en el cuadro anterior. Si aprovechamos dicha disposición (la mitad positivos y la otra mitad negativos), para formar dos bloques de 8 tratamientos cada uno, de esta manera los grupos que formen el primer bloque realizarán la prueba durante la mañana, y los que pertenezcan al segundo, durante la tarde. Así, haríamos confuso el efecto de la interacción cuádruple, con la variable "mañana y tarde". Por otra parte, hemos de tener en cuenta que la asignación de los bloques a la mañana o tarde será al azar.

3. RESUMEN

Hemos intentado presentar la problemática que entraña la utilización del diseño experimental para la ciencia de la conducta. Si bien, en principio, todo plan experimental encierra ya de por sí un determinado diseño, no por ello debemos ignorar la existencia de modelos de diseños cuya utilidad y eficacia se ha demostrado a lo largo de estos últimos años.

Nuestra intención no ha sido la de desarrollar exhaustivamente la rica temática que implica la utilización del diseño, solamente hemos pretendido presentar un cuadro básico de los principales diseños que a nuestro juicio pueden ser empleados con éxito en la investigación científica.

Hemos de señalar, por otra parte, que si al principio hemos establecido la división clásica del diseño en función de la cantidad de variables experimentales (diseños simples y factoriales), hemos considerado de mucha más utilidad presentar una clasificación triple (diseños bivalentes, multivalentes y factoriales), porque creemos que dicha división refleja más de cerca la utilización del diseño en el campo de la conducta desde una preceptiva histórica. A medida que las ciencias de la conducta se han desarrollado se ha operado un creciente cambio en su uso. Y si inicialmente podemos comprobar que la mayoría de los diseños eran de tipo de "dos grupos", cada vez es mayor el empleo de diseños más complejos.

El deseo de presentar un esquema básico de los principales diseños aplicables a nuestra ciencia, ha sido uno de los propósitos básicos de este escrito. El que le interese profundizar en alguno de ellos, puede utilizar la bibliografía que a continuación presentamos.

4. BIBLIOGRAFÍA

- CAMPBELL y STANLEY: *Experimental and quasi-experimental designs for research*. Rand McNally and Comp., Chicago, 1963.
- COCHRAN, W. G., y COX, G. M.: *Diseños experimentales*. Trillas, Méjico, 1965 (original inglés, 1957).
- COX, D. R.: *Planning of experiments*. J. Wiley, Nueva York, 1958.

- EDWARDS, A. L.: *Experimental design in psychological research*, 3.^a ed., Rinehart, Nueva York, 1968.
- FINNEY, D. J.: *Experimental design and its statistical basis*. Univ. Chicago Press, Chicago, 1955.
- FINNEY, D. J.: *An introduction to the theory of experimental design*. Univ. Chicago Press, Chicago, 1960.
- FISHER, R. A.: *The design of experiments*. Hafner, Nueva York, 1953.
- LI, CHING CHUN: *Introducción a la estadística experimental*. Omega, Barcelona, 1969.
- LINDQUIST, E. F.: *Design and analysis of experiments in psychology and education*. Houghton Mifflin Comp., Boston, 1953.
- KIMMEL, H. D.: *Experimental principles and design in psychology*. Ronald Press Comp., Nueva York 1970.
- RAY, W. S.: *An introduction to experimental design*. McMillan, Nueva York, 1960.
- REUCHLIN, M.: "Utilisation in psychologie de certains plans d'expérience", *Ann. Psychol.*, 53, 58-81, 1953.
- UNDERWOOD, B. J.: *Psicología experimental*. Trillas, Méjico, 1972.
- WINER, B. J.: *Statistical principles in experimental design*, 2.^a ed. McGraw-Hill, Nueva York, 1971.

MODELOS MATEMÁTICOS DEL ANÁLISIS DE LA VARIANZA

por CARLOS CUADRAS

1. Matriz de un diseño factorial

Sea Y una variable observada en condiciones distintas, de acuerdo con un diseño factorial. El modelo matemático para un diseño, desde un punto de vista general, consta de:

1) m parámetros desconocidos,

$$\beta_1, \beta_2, \dots, \beta_m$$

que deben ser estimados a partir de un muestreo de tamaño n .

2) Una matriz de elementos conocidos,

$$X = \begin{bmatrix} x_{11} & \dots & x_{1m} \\ \dots & \dots & \dots \\ x_{n1} & \dots & x_{nm} \end{bmatrix}$$

llamada matriz del diseño factorial.

3) Unas variables concomitantes asociadas a cada observación,

$$c_{i1} \ c_{i2} \ \dots \ c_{iq} \quad i = 1, \dots, n$$

que tienen una influencia externa sobre el diseño factorial, en relación con q parámetros desconocidos: $\gamma_1, \dots, \gamma_q$.

El modelo matemático, supone que cada observación de Y :

$$y_1, y_2, \dots, y_n$$

depende de estos parámetros según la relación:

a) $y_i = x_{i1} \beta_1 + \dots + x_{im} \beta_m + e_i$
si no hay variables concomitantes.

b) $y_i = x_{i1} \beta_1 + \dots + x_{im} \beta_m + c_{i1} \gamma_1 + \dots + c_{im} \gamma_m + e_i$
en caso de haberlas.

De esta relación, la combinación de parámetros corresponde al *valor medio* de la variable. Los valores: e_1, e_2, \dots, e_n , son aleatorios, independientes, con distribución normal $N(0, \sigma)$.

(La desviación típica es común para todos ellos.)

Una *hipótesis estadística* se expresa mediante una formulación acerca de los parámetros, que, en general, adopta la expresión:

$$H_0 \quad \begin{cases} h_{11} \beta_1 + \dots + h_{1m} \beta_m = 0 \\ h_{k1} \beta_1 + \dots + h_{km} \beta_m = 0 \end{cases}$$

Corresponde a la inferencia estadística construir el test adecuado que permita decidir (con un cierto nivel de significación) si debe aceptarse o rechazarse la hipótesis. Una hipótesis estadística puede demostrarse si los vectores (h_{11}, \dots, h_{im}) son combinación lineal de las filas de la matriz de diseño X . En caso contrario, no es posible efectuar un test estadístico para decidir la validez de la hipótesis.

2. Diseños antes y después

Se asigna al azar un grupo de $n = n_1 + n_2$ individuos en dos grupos: E (grupo experimental) y C (grupo control). Una variable Y (la medida de un determinado carácter), se observa antes y después del tratamiento.

Grupo	Observaciones antes	Observaciones después
E	$c_{11} \dots c_{1n_1}$	$y_{11} \dots y_{1n_1}$
C	$c_{21} \dots c_{2n_2}$	$y_{21} \dots y_{2n_2}$

El *modelo factorial* contiene cuatro parámetros.

Se considera que el valor de Y , obtenido antes, tiene un efecto *concomitante*.

Grupo	Valor medio	Media general	Efecto grupo	Efecto concomitante
E	$E(y_{1i})$	$= \mu$	$+ \alpha_1$	$+ c_{1i} \cdot \gamma$
C	$E(y_{2i})$	$= \mu$	$+ \alpha_2$	$+ c_{2i} \cdot \gamma$

La comparación de grupo E con grupo C, se formula mediante la hipótesis estadística:

$$H_0 : \alpha_1 = \alpha_2 = 0$$

La matriz del diseño es:

$$X = \begin{pmatrix} \mu & \alpha_1 & \alpha_2 & \gamma \\ \hline 1 & 1 & 0 & c_{11} \\ \hline 1 & 1 & 0 & c_{1^{n_1}} \\ \hline 1 & 0 & 1 & c_{21} \\ \hline 1 & 0 & 1 & c_{2^{n_2}} \end{pmatrix}$$

También puede interesar comprobar si el efecto concomitante de los valores iniciales es nulo.

$$H_0: \gamma = 0$$

3. Diseño de dos grupos al azar

Se asigna, al azar, una muestra de $n = n_1 + n_2$ individuos en dos grupos: E (experimental), C (control). Una variable Y se observa después de un tratamiento. Las observaciones serán:

Grupo	Observaciones después
E	$y_{11} \dots \dots \dots y_{1^{n_1}}$
C	$y_{21} \dots \dots \dots y_{2^{n_2}}$

El modelo factorial depende de tres parámetros:

Grupo	Valor medio		Media general		Efecto grupo
E	$E(y_{1i})$	=	μ	+	α_1
C	$E(y_{2i})$	=	μ	+	α_2

Habrà diferencia entre los dos grupos si se ha podido rechazar la hipótesis estadística:

$$H_0: \alpha_1 = \alpha_2 = 0$$

4. Diseños intrasujetos con tratamientos contrabalanceados

Se administra a un solo sujeto, dos tratamientos, uno después de otro. Se le vuelven a administrar los dos tratamientos, pero en orden inverso. Si la segunda vez se ha hecho en condiciones tales que quede eliminado el efecto del primer tratamiento, el modelo factorial dependería de cuatro parámetros.

Tratamiento	Valor medio		Media general		Efecto tratamiento	Interacción	
1	$E(y_1)$	=	μ	+	α_1		
2	$E(y_2)$	=	μ	+	α_2	+	γ
3	$E(y_3)$	=	μ	+	α_2		
4	$E(y_4)$	=	μ	+	α_1	+	γ

Se debe comprobar la hipótesis

$$H_0 \quad \alpha_1 = \alpha_2 = 0$$

Se impone la restricción $\alpha_1 + \alpha_2 = 0$, y entonces, indicando $\alpha_1 = \alpha$, $\alpha_2 = -\alpha$, la matriz del diseño es

$$X = \begin{pmatrix} \mu & \alpha & \gamma \\ 1 & 1 & 0 \\ 1 & -1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Obsérvese que la hipótesis $H_0 \quad \alpha = 0$ es demostrable porque el vector $(0 \ 1 \ 0)$ es combinación lineal de las filas de X .

5. Diseño de grupos completamente aleatorizados

Se asignan al azar en K grupos, un total de $n = n_1 + \dots + n_k$ individuos y se toman medidas de una variable Y (por ejemplo, después de un tratamiento, en las condiciones del grupo que corresponda).

Indicando:

$$y_{i1} \dots y_{in}$$

las observaciones del grupo i , el diseño factorial depende de $k+1$ parámetros con la restricción:

$$\sum_{i=1}^k \alpha_i = 0$$

Grupo	Valor medio		Media general		Efecto grupo
n.º i	$E(y_{i,h})$	=	μ	+	α_i

El test de comparación de grupos, se expresa por la hipótesis nula:

$$H_0 \quad \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$$

6. Diseño de bloques aleatorizados

Los individuos se organizan en bloques, de forma que los individuos dentro de un mismo bloque sean lo más homogéneos posible. Si se desean comparar entonces K grupos, en relación a una variable Y , se asignan, dentro de cada bloque, totalmente al azar los K grupos. Las observaciones son:

	Grupo 1	Grupo 2	Grupo K
Bloque 1	y_{11}	y_{12}	y_{1k}
Bloque S	y_{s1}	y_{s2}	y_{sk}

Puede haber más de una observación, para un grupo y un mismo bloque. El diseño depende de $(k + s + 1)$ parámetros.

Grupo — Bloque	Valor medio	Media general	Efecto grupo	Efecto bloque
Grupo i — Bloque j	$E_{(ij)}$	$= \mu$	$+ \alpha_i$	$+ \beta_j$

Hay dos tipos de hipótesis referentes a los parámetros:

Comparación de grupos: $H_0 \alpha_1 = \alpha_2 = \dots = \alpha_s = 0$

Comparación de bloques: $H_0 \beta_1 = \beta_2 = \dots = \beta_s = 0$

7. Cuadrados latinos

Si los individuos o unidades experimentales dependen de dos factores (causas de variación), y deben asignarse varios tratamientos, entonces el experimento requeriría un número elevado de observaciones, las que resultan de combinar los niveles de los factores con todos los tratamientos. El diseño de cuadrados latinos es un artificio para ahorrar observaciones. Consiste en combinar los dos factores (lo que nos da una disposición en filas y columnas) y asignar los tratamientos de forma que cada uno aparezca una sola vez en cada fila y cada columna. Una observación de una medida Y , se indicará en la forma:

$y_{ij}(k)$ i — nivel primer factor
 j — nivel segundo factor
 k — tratamiento asignado.

Debe observarse que el tratamiento k depende de i, j .

El diseño factorial es:

Valor medio		Media general		Efecto 1.º factor		Efecto 2.º factor		Efecto tratamiento
$E(y_{ij}(k))$	=	μ	+	α_i	+	β_j	+	γ_k

Interesa comparar los tratamientos, mediante un test que decida si debe aceptarse o rechazarse la hipótesis:

$$H_0 \quad \gamma_1 = \gamma_2 = \dots = \gamma_m = 0 \quad (m \text{ tratamientos})$$

8. Diseño de dos factores con interacción

Si la causa de variación sobre una variable Y , son dos factores, con r y s niveles, cabe tener en cuenta los efectos principales de cada factor, y la interacción o influencia de la relación entre los niveles de uno y otro factor. Las observaciones se obtienen según el siguiente esquema (dos factores A y B , con tres y dos niveles):

		Factor A		
		1	2	3
Factor B	1	y_{111}	y_{211}	y_{311}
		.	.	.
		.	.	.
	2	y_{11n}	y_{21n}	y_{31n}
		y_{121}	y_{221}	y_{321}
		.	.	.
2	.	.	.	
	.	.	.	
	y_{12n}	y_{22n}	y_{32n}	

Habrán tres clases de parámetros. El modelo es:

Valor medio	Media general		Efecto factor A		Efecto factor B		Interacción
$E(y_{ijk})$	μ	+	α_i	+	β_j	+	γ_{ij}

Es importante testar la significación de los efectos principales y las interacciones.

Efecto de A	$H_0 \alpha_1 = \dots = \alpha_r = 0$
Efecto de B	$H_0 \beta_1 = \dots = \beta_s = 0$
Interacciones	$H_0 \gamma_{11} = \dots = \gamma_{r1} = \dots = \gamma_{rs} = 0$

BIBLIOGRAFÍA

1. CHIN CHUN LI: *Introducción a la estadística experimental*. Ed. Omega, Barcelona, 1969.
2. COCHRAN, W. G., y COX, G. M.: *Experimental Designs*. Ed. Wiley, Nueva York, 1966.
3. SCHEFFE, H.: *The Analysis of Variance*. Ed. Wiley, Nueva York, 1959.

TÉCNICAS DE MUESTREO

por PEDRO BATALLE

1. INTRODUCCIÓN

1.1. *Ventajas del Método de Muestreo*

En primer lugar, dirigido en especial a todos aquellos que se resisten a emplear las muestras en sus estudios, y a sospechar de ellas, voy a citar las principales ventajas del muestreo, comparado con la enumeración completa:

Costo reducido.

Si los datos se obtienen a partir de una pequeña fracción del total, los gastos son menores que llevando a cabo un censo completo.

Mayor rapidez.

Por la misma razón, los datos pueden ser recolectados más rápidamente con una muestra que con una enumeración completa.

Mayor exactitud.

Si el volumen de trabajo es reducido se puede emplear personal capacitado y se puede realizar una supervisión cuidadosa del trabajo de campo y del procesamiento de los resultados, redundando todo ello en una mayor exactitud.

1.2. *Principales etapas en una encuesta por muestreo*

Como introducción al papel de la teoría del muestreo, voy a describir los principales pasos existentes en la planeación y ejecución de una encuesta por muestreo.

1) *Objetivos de la encuesta.*

Es imprescindible una exposición clara de los principales objetivos.

2) *Población bajo muestreo.*

La población es el conjunto del que se elige la muestra. Deben existir unas reglas, previamente establecidas, que nos permitan discernir en todos los casos si un individuo pertenece, o no, a la población estudiada.

La población que va a ser muestreada (población muestreada) debe coincidir con la población acerca de la que se desea hacer inferencia (población objeto). De no ser así, las conclusiones sacadas de la muestra, sólo se pueden aplicar a la población muestreada.

3) Datos que deben recogerse.

Se debe comprobar que no se ha omitido ningún dato esencial. Existe la tendencia de preguntar demasiadas cosas, que luego no serán analizadas.

Hay que tener en cuenta que un cuestionario demasiado largo baja la calidad de las respuestas.

4) Nivel de precisión deseado.

Los resultados están sujetos a una cierta incertidumbre, debido a que sólo ha sido investigada una parte de la población y debido a los errores de medición. Esta incertidumbre puede reducirse tomando muestras mayores. Desgraciadamente esto cuesta tiempo y dinero; en consecuencia es un paso importante la especificación del nivel de precisión deseado. Es decir, pensar en la cantidad de error que puede tolerarse en los estimadores, compatible con la toma de buenas decisiones.

5) Métodos de medición.

Se debe escoger el método de medición e inspección de la población. Existe también un trabajo preliminar que es la elaboración del cuestionario.

6) El marco.

Antes de la elección de la muestra, la población debe ser dividida en partes llamadas "unidades de muestreo" o "unidades". Estas unidades deben cubrir la totalidad de la población y no deben superponerse.

La construcción de esta lista de unidades, llamada marco, es, frecuentemente, uno de los principales problemas prácticos.

7) Elección de la muestra.

Existen actualmente muchos planes para seleccionar una muestra. Para cada plan pueden hacerse estimaciones del tamaño de la muestra partiendo del nivel de precisión deseado.

Antes de tomar una decisión debe también tenerse en cuenta el coste y el tiempo involucrados en cada plan.

8) La encuesta piloto.

Es muy útil el probar el cuestionario y los métodos de campo en pequeña escala. Acostumbra a producir mejoras en el cuestionario y puede revelar algunos problemas existentes.

9) Organización del trabajo de campo.

El personal debe recibir un entrenamiento en relación con el propósito de la encuesta y los métodos de medición a emplear. Deberá supervisarse su trabajo. Deben hacerse planes para manejar las no-respuestas.

10) Resumen y análisis de los datos.

Se debe decidir acerca del procedimiento de tabulación y sobre los distintos métodos de estimación de la misma información. En la presentación de los resultados es aconsejable indicar el error esperado en los estimadores.

11) Información ganada para futuras encuestas.

Cualquier muestra terminada es potencialmente una guía para mejorar el futuro muestreo. Además, los errores de ejecución que han habido, se procurará que no ocurran en la siguiente encuesta.

1.3. *Importancia de la Teoría del Muestreo*

El objeto de la lista de pasos anterior, es el de hacer énfasis sobre los aspectos prácticos del muestreo. En algunos de estos pasos, la teoría del muestreo tiene una importancia secundaria. Ahora bien, todas las fases son importantes; una deficiencia en una de ellas puede arruinar todo el trabajo.

El propósito de la teoría del muestreo es el de desarrollar métodos de selección de muestras y métodos de estimación que proporcionen estimadores que sean lo suficientemente precisos.

La precisión de un procedimiento de muestreo se juzga examinando la distribución de frecuencias generada para el estimador, en el caso de que el procedimiento fuese aplicado una y otra vez a la misma población.

Aquí, se puede introducir una simplificación. Con muestras un poco grandes, existe una buena razón, para suponer que los estimadores se distribuyen siguiendo una ley normal. Cuando un estimador se distribuye normalmente, la forma general de la distribución de frecuencias es conocida si sabemos la media y la desviación tipo. Una parte considerable de la teoría del muestreo se dedica a encontrar las fórmulas para estas medias y variancias.

La teoría del muestreo diferencia aquellos casos en los que la población tiene un número finito de unidades. Entonces, los resultados a los que se llega son algo más complicados que los correspondientes a poblaciones infinitas.

Ahora bien, en todos aquellos casos en que el tamaño de la muestra es pequeño respecto al de la población, los resultados derivados de una población infinita son totalmente adecuados.

1.4. *Muestreo probabilístico*

Todos los procedimientos de muestreo tienen en común las siguientes propiedades:

1) Es posible definir un conjunto de muestras M_1, M_2, \dots, M_n , mediante la aplicación de un procedimiento a una población específica que conduzca a la selección de estas muestras.

2) A cada posible muestra M_i le ha sido asignada una probabilidad de selección p_i .

3) Seleccionamos una de las M_i por un proceso mediante el cual cada M_i tiene una probabilidad p_i de ser elegida.

4) Debe establecerse el método para calcular el estimador de la muestra, que debe conducir a un estimador único para cualquier muestra específica.

Entonces podremos calcular la distribución de frecuencias de los estimadores que se generen si el método es aplicado repetidamente a la misma población.

El término de "muestreo probabilístico" se refiere a un procedimiento de este tipo. Desde luego que no es el único camino para sacar una muestra.

En la práctica rara vez obtenemos una muestra basada en probabilidades estableciendo las M_i y las p_i , como se dijo antes. Esto es imposible de realizar con una población grande, en donde un procedimiento de muestreo puede producir miles de millones de posibles muestras.

La obtención de la muestra se acostumbra a hacer especificando las probabilidades de las unidades individuales y seleccionando dichas unidades una a una o en grupos, hasta que se tiene la muestra deseada.

1.5. Uso de la distribución normal

Las muestras son a menudo lo suficientemente grandes para que el estimador obtenido en ellas se distribuya de forma normal.

Un estimador $\bar{\mu}$ obtenido por un procedimiento de muestreo se llama estimador insesgado de alguna característica de la población μ , si el valor medio de $\bar{\mu}$, calculado con todas las muestras posibles, es igual a μ . Es decir:

$$E(\bar{\mu}) = \sum_{i=1}^n p_i \bar{\mu}_i = \mu$$

Supongamos que, en una muestra, hemos calculado el estimador $\bar{\mu}$ y su desviación tipo $\sigma_{\bar{\mu}}$ (error tipo). ¿Es bueno este estimador? No podemos saber el valor exacto del error del estimador ($\bar{\mu} - \mu$), pero por las propiedades de la curva normal, sabemos que:

— la probabilidad de que el error absoluto ($\bar{\mu} - \mu$) sea mayor que $\sigma_{\bar{\mu}}$ es 0,32;

— la probabilidad de que el error absoluto ($\bar{\mu} - \mu$) sea mayor que $1,96 \sigma_{\bar{\mu}} = 2 \sigma_{\bar{\mu}}$ es 0,05;

— la probabilidad de que el error absoluto ($\bar{\mu} - \mu$) sea mayor que $2,58 \sigma_{\bar{\mu}}$ es 0,01.

A los límites $\bar{\mu} \pm z_{\alpha} \cdot \sigma_{\bar{\mu}}$ se les llama límites de confianza inferior y superior.

Cuando la muestra es pequeña, las tablas de la distribución t de Student-Fisher, sustituyen a las normales en el cálculo de los límites de confianza:

$$\bar{\mu} \pm t_{\alpha} \cdot \sigma_{\bar{\mu}}$$

La tabla normal con la de t casi no tiene diferencia, si el número de grados de libertad excede a 60.

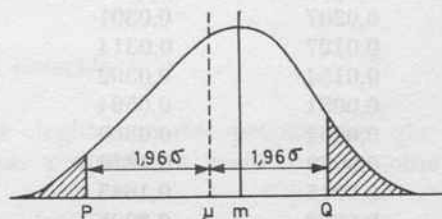
1.6. Sesgo y sus efectos

Es importante estudiar el sesgo por dos razones:

- 1) En algunos de los problemas más comunes, los estimadores que son más convenientes son sesgados.
- 2) Aun con estimadores que son insesgados, los errores de medición y las no respuestas pueden producir sesgos, en los números que calculamos a partir de los datos.

Voy a examinar el efecto del sesgo.

Supongamos que el estimador \bar{u} está normalmente distribuido alrededor de una media n que está a una distancia B del verdadero valor μ de la población. La magnitud del sesgo es $B = m - \mu$.



Supongamos que se conoce la existencia de un sesgo. Calculamos la desviación tipo σ de la distribución del estimador (será la desviación tipo respecto la media de la distribución, no respecto la verdadera media μ). Estamos usando σ y no $\sigma\mu$.

Si no hubiera sesgo, podríamos decir que la probabilidad de que el error fuese mayor que $1,96\sigma$ es 0,05.

Veamos cómo la presencia del sesgo distorsiona esta probabilidad.

Examinemos los dos lados de la distribución separadamente.

Para el extremo derecho, la probabilidad de un error mayor que $1,96\sigma$ es el área rayada a partir de Q .

Este límite inferior Q sería:

$$\frac{\mu - m}{\sigma} + 1,96 = 1,96 - \frac{B}{\sigma}$$

Para el otro extremo la probabilidad de un error mayor que $1,96\sigma$ sería el área rayada a partir de P .

Dicho punto P sería:

$$-1,96 - \frac{B}{\sigma}$$

Es decir, que la perturbación depende únicamente de la proporción de sesgo respecto la desviación tipo.

A continuación se da una tabla que expresa la variación experimentada en la probabilidad de que el error sea mayor que $1,96 \sigma$, en función de distintos valores de la proporción $\frac{B}{\sigma}$.

B/σ	Probabilidad de error		Total
	$< -1,96 \sigma$	$> 1,96 \sigma$	
0,02	0,0238	0,0262	0,0500
0,04	0,0228	0,0274	0,0502
0,06	0,0217	0,0287	0,0504
0,08	0,0207	0,0301	0,0508
0,10	0,0197	0,0314	0,0511
0,20	0,0154	0,0392	0,0546
0,40	0,0091	0,0594	0,0685
0,60	0,0052	0,0869	0,0921
0,80	0,0029	0,1230	0,1259
1,00	0,0015	0,1685	0,1700
1,50	0,0003	0,3228	0,3231

Observemos que el sesgo tiene poca influencia, siempre que sea inferior a una décima de la desviación tipo.

Para $B = \sigma$ la probabilidad de error es de 0,17, más de tres veces el valor supuesto.

Observemos, por último, que los dos lados de la distribución se ven afectados de forma distinta.

2. MUESTREO SIMPLE ALEATORIO

2.1. Muestreo simple aleatorio

Supongamos una población que contiene un número N de unidades. Si todas estas unidades pueden distinguirse una de la otra, el número de muestras distintas de tamaño n que pueden sacarse de N unidades es:

$$\binom{N}{n} = C_N^n = \frac{N!}{n! (N-n)!}$$

Por ejemplo si la población contiene cinco unidades identificadas por A, B, C, D, E, existen diez muestras diferentes de tamaño tres, que son:

$A B C \quad A B D \quad A B E \quad A C D \quad A C E$
 $A D E \quad B C D \quad B C E \quad B D E \quad C D E$

El muestreo simple aleatorio es un método de selección de n unidades sacadas de N , de tal manera que cada una de las muestras tiene la misma posibilidad u oportunidad de ser escogida.

En la práctica, una muestra simple al azar es sacada unidad por unidad. Las unidades en la población son numeradas de 1 a N . A continuación se sacan series de números al azar entre 1 y N , ya sea por medio de una tabla de números al azar o colocando los números 1 a N en una urna, mezclándolos perfectamente, hasta completar los n de la muestra.

El muestreo se considera sin reemplazo o reposición, es decir, que un número sacado de la urna no es repuesto, o utilizando la tabla de números aleatorios, un número sacado previamente es ignorado.

2.2. Definiciones y notación

En una muestra elegimos ciertas propiedades que intentamos medir en cada unidad. Dichas propiedades reciben el nombre de características o atributos.

Las letras mayúsculas se refieren a las características de la población y las minúsculas a las de la muestra. Para los totales y las medias tenemos las siguientes definiciones:

	Población	Muestra
Total	$Y = \sum_{i=1}^N y_i = y_1 + y_2 + \dots + y_N$	$\sum_{i=1}^n y_i = y_1 + y_2 + \dots + y_n$
Media	$\bar{Y} = \frac{y_1 + y_2 + \dots + y_N}{N} = \frac{\sum_{i=1}^N y_i}{N}$	$\bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n} = \frac{\sum_{i=1}^n y_i}{n}$

El interés del muestreo se centra, frecuentemente, en las siguientes cuatro características de la población:

1. Media = \bar{Y}
2. Total = Y
3. Proporción de dos totales o medias $R = Y/X = \bar{Y}/\bar{X}$
4. Proporción de unidades que caen dentro de alguna clase definida.

El símbolo $\hat{\Lambda}$ identifica a un estimador muestral de una característica de la población. Los estimadores más simples son:

Media de la población \bar{Y}	$\bar{Y} = \bar{y} =$ media muestral
Total de la población Y	$\hat{Y} = \bar{y} = N \sum y_i/n$
Proporción de la población R	$\hat{R} = \bar{y}/\bar{x} = \sum y_i/\sum x_i$

En el estimador de \hat{Y} el factor N/n , se le llama factor de expansión o inflación. Su inverso n/N , proporción del tamaño de la muestra respecto la población, se le llama fracción de muestreo y se identifica por la letra f .

2.3. Propiedades de los estimadores

Recordemos que un método de estimación se llama consistente si el estimador se vuelve exactamente igual al valor de la población cuando $n = N$.

Un método de estimación es insesgado si el valor promedio del estimador, tomado sobre todas las muestras posibles de tamaño n , es exactamente igual al verdadero valor de la población.

Teorema: La media muestral \bar{y} es un estimador insesgado de \bar{Y} .

Demostración:

$$\text{Por definición } E \bar{y} = \frac{\sum \bar{y}}{C_N^n} = \frac{\sum (y_1 + y_2 + \dots + y_n)}{n[N!/n! (N-n)!]}$$

Para evaluar esta suma, que se extiende a todas las C_N^n muestras, buscamos en cuántas muestras aparece un valor específico cualquiera y_i . Como están disponibles otras $(N-1)$ unidades para el resto de la muestra y otras $(n-1)$ unidades para completar la muestra, el número de muestras conteniendo y_i es:

$$C_{N-1}^{n-1} = \frac{(N-1)!}{(n-1)! (N-n)!}$$

De donde:

$$\sum (y_1 + y_2 + \dots + y_n) = \frac{(N-1)!}{(n-1)! (N-n)!} (y_1 + y_2 + \dots + y_n)$$

Sustituyendo tenemos:

$$\begin{aligned} E \bar{y} &= \frac{(N-1)!}{(n-1)! (N-n)!} \frac{n! (N-n)!}{n \cdot N!} (y_1 + y_2 + \dots + y_n) = \\ &= \frac{y_1 + y_2 + \dots + y_n}{N} = \bar{Y} \end{aligned}$$

Corolario. $\hat{Y} = N \bar{y}$ es un estimador insesgado de la población total Y .

2.4. Varianzas de los estimadores

La varianza de y_i en una población finita viene definida por:

$$\sigma^2 = \frac{\sum_1^N (y_i - \bar{Y})^2}{N}$$

A veces esta expresión se representa así: $S^2 = \frac{\sum_1^N (y_i - \bar{Y})^2}{N - 1}$

Esta convención la usan aquellos que enfocan la teoría del muestreo por medio del análisis de varianza.

Consideremos ahora la varianza de \bar{y} , es decir $E(\bar{y} - \bar{Y})^2$ tomada sobre todas las C_N^n muestras.

Se puede demostrar que la varianza de la media \bar{y} de una muestra simple aleatoria es:

$$V(\bar{y}) = E(\bar{y} - \bar{Y})^2 = \frac{S^2}{n} \frac{N - n}{N} = \frac{S^2}{n} (1 - f)$$

en donde $f = n/N$ es la fracción de muestreo.

Dicho teorema tiene varios corolarios.

Corolario 1) El error tipo de \bar{y} es:

$$\sigma_{\bar{y}} = \frac{S}{\sqrt{n}} \sqrt{(N - n)/N} = \frac{S}{\sqrt{n}} \sqrt{1 - f}$$

Corolario 2) La varianza de $\hat{Y} = N \bar{y}$, como un estimador de la población total Y es:

$$V(\hat{Y}) = E(\hat{Y} - Y)^2 = \frac{N^2 S^2}{n} \cdot \frac{(N - n)}{N} = \frac{N^2 S^2}{n} (1 - f)$$

Corolario 3) El error tipo de \hat{Y} es:

$$\sigma_{\hat{Y}} = \frac{NS}{\sqrt{n}} \sqrt{(N - n)/N} = \frac{NS}{\sqrt{n}} \sqrt{1 - f}$$

2.5. La corrección por población finita

Para una muestra al azar de tamaño n , obtenida de una población infinita, la varianza de la media es σ^2/n . El único cambio en este resultado, cuando la

población es finita, es la introducción del factor $(N - n)/N$. Los factores $(N - n)/N$ para la varianza, y $\sqrt{(N - n)/N}$ para el error tipo, se donominan correcciones debidas a población finita. Estas correcciones están dadas por algunos autores con un divisor $(N - 1)$ en lugar de N .

Siempre y cuando la fracción n/N sea pequeña, estos factores toman valores cercanos a la unidad y el tamaño de la población como tal no tiene un efecto directo en el error tipo de la media de la muestra.

En la práctica, el factor de corrección por población finita puede ignorarse siempre que la fracción de muestreo no exceda del 5% y en algunos casos del 10%. El efecto de ignorar la corrección es sobrestimar el error tipo del estimador \bar{y} .

2.6. Estimación del error tipo de una muestra

Las fórmulas para los errores tipo del estimador de una media y total de la población, se usan principalmente para tres propósitos:

1) para comparar la precisión obtenida por el muestreo simple aleatorio con otros métodos de muestreo.

2) para estimar el tamaño de la muestra que se necesita en una encuesta que está siendo planeada.

3) para estimar la precisión realmente obtenida en una encuesta que se haya terminado.

Las fórmulas involucras a S^2 , la varianza de la población. Ésta no será conocida en la práctica, pero puede ser estimada a partir de los datos de la muestra.

Un resultado importante es el que establece el siguiente teorema.

Teorema. En una muestra simple al azar.

$$s^2 = \frac{\sum_1^n (y_i - \bar{y})^2}{n - 1}$$

es un estimador insesgado de

$$S^2 = \frac{\sum_1^N (y_i - \bar{Y})^2}{N - 1}$$

Corolario. Los estimadores insesgados de la varianza de:

\bar{y} y $\bar{Y} = \bar{y}$ son:

$$v(\bar{y}) = \frac{s^2}{y} = \frac{s^2}{n} \left(\frac{N - n}{N} \right) = \frac{s^2}{n} (1 - f)$$

$$v(\bar{Y}) = \frac{s^2}{y} = \frac{N^2 s^2}{n} \left(\frac{N - n}{N} \right) = \frac{N^2 s^2}{n} (1 - f)$$

Para los errores tipo, tomamos:

$$s\hat{y} = \frac{s}{\sqrt{n}} \cdot \sqrt{1-f}, \quad s\hat{t} = \frac{Ns}{\sqrt{n}} \sqrt{1-f}$$

Estos estimadores son ligeramente sesgados, pero en la mayoría de aplicaciones este sesgo no tiene importancia.

2.7. Límites de confianza

Generalmente se presupone que los estimadores \bar{y} y \hat{Y} se distribuyen en forma normal respecto al valor paramétrico correspondiente a la población.

Entonces los límites de confianza superior e inferior para la media y total de la población son:

Media

$$\hat{Y}_{\text{inf}} = \bar{y} - \frac{Zs}{\sqrt{n}} \sqrt{1-f} \quad \hat{Y}_{\text{sup}} = \bar{y} + \frac{Zs}{\sqrt{n}} \sqrt{1-f}$$

Total

$$\hat{Y}_{\text{inf}} = N\bar{y} - \frac{ZNs}{\sqrt{n}} \sqrt{1-f} \quad \hat{Y}_{\text{sup}} = N\bar{y} + \frac{ZNs}{\sqrt{n}} \sqrt{1-f}$$

El símbolo Z es la puntuación tipo de la ley normal, es decir, el valor del desvío normal correspondiente al nivel de confianza deseado.

Si el tamaño de la muestra es menor de 60, entonces se utilizan los valores correspondientes a la t de Student-Fisher, con $(n-1)$ grados de libertad, correspondientes a la varianza estimada s^2 .

2.8. Estimación de una proporción o razón

La cantidad que ahora va a estimarse es la proporción de dos variables, ambas variando de unidad en unidad.

El parámetro de la población a estimar es:

$$R = \frac{\sum_1^N y_i}{\sum_1^N x_i}$$

El estimador correspondiente derivado de una muestra es:

$$\hat{R} = \frac{\sum_1^n y_i}{\sum_1^n x_i} = \frac{\bar{y}}{\bar{x}}$$

La distribución muestral de \hat{R} es más complicada que la de \bar{y} , porque ambos, el numerador \bar{y} y el denominador \bar{x} varían de muestra a muestra.

En muestras pequeñas la distribución de \hat{R} es asimétrica y \hat{R} acostumbra a ser un estimador sesgado de R .

En muestras grandes la distribución de R tiende a la normalidad y el sesgo se vuelve despreciable.

Teorema.

Si las variables y_i, x_i son medidas en cada unidad de una muestra simple aleatoria de tamaño n , supuestamente grande, la varianza de $\hat{R} = \bar{y}/\bar{x}$ es, aproximadamente:

$$V(\hat{R}) = \frac{1-f}{n\bar{x}^2} \frac{\sum_{i=1}^N (y_i - R x_i)^2}{N-1}$$

en donde $R = \bar{Y}/\bar{X}$ es la proporción de las medias de la población y $f = n/N$.

Como un estimador muestral de:

$$\frac{\sum_{i=1}^N (y_i - R x_i)^2}{N-1}$$

se acostumbra a tomar:

$$\frac{\sum_{i=1}^n (y_i - \hat{R} x_i)^2}{n-1}$$

Se puede demostrar que este estimador tiene un sesgo $1/n$. Para el error tipo estimado de R , se tiene:

$$s(R) = \frac{\sqrt{1-f}}{\sqrt{n}\bar{x}} \sqrt{\frac{\sum (y_i - R x_i)^2}{n-1}}$$

2.9. Validez de la aproximación normal

Se han hecho muchos estudios en teoría de probabilidades sobre la distribución de las medias de muestras aleatorias. Se ha probado que, para cualquier población que tiene una desviación tipo finita, la distribución de la media muestral tiende a la normalidad conforme n aumenta. (Estudios de Feller, 1957, con poblaciones infinitas.)

Para muestreo sin reemplazo con poblaciones finitas, Hajek (1960) dio las condiciones necesarias y suficientes bajo las cuales la distribución de la media muestral tiende a la normalidad.

Pero, ¿qué tan grande debe ser n para que la aproximación normal sea suficientemente exacta?

No hay una regla general segura para establecer qué tan grande debe ser n para poder usar la aproximación normal en el cálculo del límite de confianza.

Para poblaciones en las que la desviación principal de la normalidad consiste en una asimetría marcadamente positiva, una regla burda que se ha encontrado es:

$$n > 25 G_1^2$$

en donde G_1 es la medida de asimetría de Fisher:

$$G_1 = \frac{E (y_i - \bar{Y})^3}{\sigma^3} = \frac{1}{N\sigma^3} \sum_{i=1}^N (y_i - \bar{Y})^3$$

Una buena práctica de muestreo tiende a hacer la aproximación normal más válida. La falla en la aproximación normal ocurre, principalmente, cuando la población contiene algunos individuos extremos que dominan el promedio de muestra cuando están presentes. Sin embargo, estos extremos también tienen un efecto mucho más serio al aumentar la varianza de la muestra y disminuir la precisión. En consecuencia, es interesante segregarlos y hacer planes separados para controlarlos. Esta eliminación de los individuos extremos de la población reduce la asimetría y mejora la aproximación normal.

3. MUESTREO PARA PROPORCIONES Y PORCENTAJES

3.1. Características

En algunas ocasiones deseamos estimar el número total, la proporción, o el porcentaje de unidades de una población que posee alguna característica o atributo.

Notación.

Supongamos que todas las unidades de la población caen dentro de una de dos posibles clases C y C' . La notación es como sigue:

Número de unidades en C en:		Proporción de unidades en C en:	
Población	Muestra	Población	Muestra
A	a	$P = A/N$	$p = a/n$

El estimador muestral de P es p , y el estimador muestra de A es N_p o $N \cdot a/n$.

La distribución binomial se aplica, a menudo, a estimadores como a y p . La distribución correcta en poblaciones finitas es la hipergeométrica.

3.2. Varianzas de los estimadores muestrales

Por medio de un simple artificio, se pueden aplicar los teoremas anteriores a esta situación. Para cualquier unidad de la muestra o población, se define

y_i como 1 si la unidad está en C , o como 0 si la unidad está en C' . Para estos valores y_i en la población:

$$Y = \sum_1^N y_i = A$$

$$\bar{y} = \frac{\sum_1^N y_i}{N} = \frac{A}{N} = P$$

De la misma forma, en la muestra:

$$\bar{y} = \frac{\sum_1^n y_i}{n} = \frac{a}{n} = p$$

O sea que el problema de estimar A y P es similar a la estimación del total y la media de una población, en la cual todos los valores y_i son 1 o 0.

Expresemos S^2 y 1^2 en términos de P y p .

Nótese que:

$$\sum_1^N y_i^2 = A = NP \quad , \quad \sum_1^n y_i^2 = a = n \cdot p$$

Por tanto:

$$S^2 = \frac{\sum_1^N (y_i - \bar{y})^2}{N-1} = \frac{\sum_1^N y_i^2 - N \bar{y}^2}{N-1} = \frac{1}{N-1} (NP - NP^2) = \frac{N}{N-1} P Q$$

en donde $Q = 1 - P$.

De forma similar:

$$s^2 = \frac{\sum_1^n (y_i - \bar{y})^2}{n-1} = \frac{n}{n-1} p q$$

Aplicando los teoremas anteriores llegaríamos a los siguientes resultados:

— La proporción en la muestra $p = a/n$ es un estimador insesgado de la proporción en la población $P = A/N$.

— La varianza de p es

$$V(p) = E(p - P)^2 = \frac{S^2}{n} \left(\frac{N-n}{N} \right) = \frac{P Q}{n} \left(\frac{N-n}{N-1} \right)$$

— La varianza de $\hat{A} = N_p$ es

$$V(\hat{A}) = \frac{N^2 P Q}{n} \left(\frac{N-n}{n-1} \right)$$

— Un estimador insesgado de la varianza de p , es:

$$v(p) = s_p^2 = \frac{N-n}{(n-1)N} p_q$$

Si N es muy grande con relación a n , un estimador insesgado de la varianza de p es:

$$\frac{p q}{n-1}$$

— Un estimador de la varianza de $\bar{A} = N_p$, es

$$v(\bar{A}) = s_{NP}^2 = \frac{N(N-n)}{n-1} p_q$$

3.3. La distribución binomial

Puesto que la población es de un tipo particularmente simple, en donde la y , toma los valores 1 o 0, podemos encontrar la distribución real de frecuencias del estimador p y no simplemente su media o su varianza.

La población contine A unidades que están en la clase C y $N-A$ unidades que están en C' , en donde $P = \frac{A}{N}$. Si la primera unidad que se analiza pertenece a C , quedarán en la población $A-1$ unidades en C y $N-A$ en C' . Por tanto, la proporción de unidades en C , después de haber obtenido la primera unidad, cambia ligeramente a $(A-1)/(N-1)$. Si la primera unidad analizada se encuentra en C' la proporción en C cambia a $A/(N-1)$. Supongamos que ignoramos estas variaciones, es decir, se presupone P constante. Esto equivale a suponer que A y $N-A$ son grandes con relación al tamaño n de la muestra.

Con esta suposición, el proceso de obtención de individuos para formar la muestra consiste en una serie de n ensayos, en cada uno de los cuales la probabilidad de que la unidad que se está analizando pertenezca a C es P . Esta situación da lugar a la distribución de frecuencias conocida como binomial.

La probabilidad de que la muestra tenga a individuos de C es:

$$Pr(a) = \frac{n!}{a! (n-a)!} P^a Q^{n-a}$$

A partir de esta expresión podemos tabular la distribución de frecuencias de a , de $p = a/n$, o del total estimado N_p .

3.4. La distribución hipergeométrica

La distribución de p puede encontrarse sin la suposición de que la población es grande en relación a la muestra. Los números de unidades de las dos clases C y C' de la población, son A y A' respectivamente. Calculemos la probabilidad de que en la muestra sean a y a' , en donde:

$$a + a' = n \quad , \quad A + A' = N$$

En el muestreo simple aleatorio cada una de las $\binom{N}{n}$ selecciones diferentes de n unidades obtenidas de N tienen igual probabilidad de ser obtenidas. Para encontrar la probabilidad deseada, contamos el número de muestras que contienen exactamente a unidades pertenecientes a C y a' a C' . El número de selecciones diferentes de a unidades entre las A que están en C es $\binom{A}{a}$, mientras que el número de selecciones diferentes de a' dentro de A' es $\binom{A'}{a'}$. Cada una de las selecciones del primer tipo se puede combinar con cualquiera del segundo para dar una muestra diferente del tipo requerido. El número total de muestras es pues:

$$\binom{A}{a} \cdot \binom{A'}{a'}$$

La probabilidad de que se obtenga una muestra de tamaño n del tipo requerido es:

$$pr(a, a'/A, A') = \frac{\binom{A}{a} \cdot \binom{A'}{a'}}{\binom{N}{n}}$$

Esta es la distribución de frecuencias de a o np , y recibe el nombre de hipergeométrica.

3.5. Límites de confianza

Aproximación normal.

Los límites de confianza para P son:

$$p \pm \left[t \sqrt{1-f} \sqrt{pq/(n-1)} + \frac{1}{2n} \right]$$

en donde $f = n/N$ y t es la puntuación tipo de la ley normal correspondiente al nivel de confianza. Es más familiar el uso de $\sqrt{p1/n}$.

El último término es una corrección que produce una mejora en la aproximación normal.

El error en la aproximación normal depende de todas las cantidades n , p , N , α .

Existen tablas que presentan reglas de trabajo para decidir cuándo la aproximación normal puede usarse.

4. ESTIMACIÓN DEL TAMAÑO DE LA MUESTRA

En el planeamiento de una encuesta por muestreo, siempre se llega a un punto en el que debe tomarse una decisión acerca del tamaño de la muestra. La decisión es importante. Una muestra demasiado grande implica un desperdicio de recursos y una muestra demasiado pequeña disminuye la utilidad de los resultados. La teoría del muestreo proporciona un planteamiento general dentro del cual puede pensarse inteligentemente en relación al problema.

4.1. Análisis del problema

Los principales pasos involucrados en la elección de un tamaño de muestra son los siguientes:

- Debe haber una indicación relativa a lo que es esperado de la muestra. Esta indicación suele estar en términos de los límites de error deseados.
- Se debe encontrar alguna ecuación que relacione n con la precisión deseada.
- Esta ecuación contendrá, como parámetros, ciertas propiedades desconocidas de la población, que deberán estimarse.
- Con frecuencia sucede que los datos se publican para ciertas subdivisiones de la población y que los límites de error deseados son establecidos para cada subdivisión. Se calcula n_i separadamente para cada subdivisión y el tamaño total n se encuentra sumando las n_i .
- Usualmente se mide más de una característica en una encuesta por muestreo. Si se especifica el grado de precisión deseado para cada característica, los cálculos llevan a una serie de valores conflictivos de n . Se debe encontrar algún método para reconciliar estos valores.
- Finalmente, debe comprobarse si el valor n es consistente con los recursos disponibles. Esto requiere una estimación del costo, trabajo, tiempo y materiales requeridos para obtener el tamaño de muestra propuesto.

4.2. Especificación de la precisión

El establecimiento de la precisión deseada suele hacerse indicando la cantidad de error que estamos dispuestos a tolerar en los estimadores muestrales.

Esta cantidad se determina a la luz de los usos que tendrán los resultados de la muestra.

4.3. Fórmula de n en el muestreo para proporciones

Las unidades se clasifican en dos clases C y C' . Se ha convenido un margen de error d en la proporción estimada p de unidades pertenecientes a la clase C y hay un pequeño riesgo α que estamos dispuestos a correr de que el error real sea mayor que d ; es decir, queremos:

$$\Pr(|p - P| \geq \alpha) = \alpha$$

Supongamos un muestreo simple aleatorio y consideremos que p se distribuye normalmente:

$$\sigma_p = \sqrt{\frac{N-n}{N-1}} \cdot \sqrt{\frac{PQ}{n}}$$

Por tanto, la fórmula que relaciona n con el grado de precisión deseado es:

$$d = t \sqrt{\frac{N-n}{N-1}} \cdot \sqrt{\frac{PQ}{n}}$$

en donde t es la abscisa del punto que separa el área α de los extremos de la distribución.

Despejando n sale:

$$n = \frac{\frac{t^2 PQ}{d^2}}{1 + \frac{1}{N} \frac{t^2 PQ}{d^2} - 1}$$

En la práctica, se utiliza un estimador adelantado p de P . Si N es grande, una primera aproximación es:

$$n_0 = \frac{t^2 p q}{d^2}$$

Calculamos primero n_0 . Si n_0/N es despreciable, n_0 es una aproximación satisfactoria de n . Si no es así, n se obtiene como sigue:

$$n = \frac{n_0}{1 + (n_0 - 1)/N} = \frac{n_0}{1 + (n_0/N)}$$

Nota. — Este documento no es original. Está extraído de algunos puntos del libro *Técnicas de Muestreo* escrito por William Cochran y publicado por CECOSA. El título original en inglés es *Sampling Techniques* editado por John Wiley.

PRUEBAS DE SIGNIFICACIÓN EN EL DISEÑO EXPERIMENTAL

por PILAR IBARROLA

Cuando un investigador estudia un determinado fenómeno, lo que hace es fijarse en una cierta variable observable que es la que caracteriza al primero y que llamaremos variable de respuesta, y lo que tratará será de averiguar qué otras variables no observables, o causas, son las que están interviniendo en el fenómeno observado, determinando los valores que va tomando la variable de respuesta; llamaremos a estas nuevas variables, variables independientes.

El investigador, basándose puramente en su lógica, seleccionará unas cuantas variables independientes que, a su juicio, piense que pueden ser las causas más importantes que pueden influir en el fenómeno, y, una vez seleccionadas, planteará un modelo matemático que le dé la relación entre la variable de respuesta y las variables independientes. ¿Pero, cómo saber que no se ha equivocado? ¿Cómo comprobar si estas variables, que él subjetivamente ha incluido en el modelo, influyen realmente o no en la variable de respuesta?

La solución a este problema está, precisamente, en las llamadas pruebas de significación.

Casi categóricamente podemos afirmar que, un modelo, para que traduzca bien la realidad, ha de ser estocástico, es decir, en él hemos de incluir una variable aleatoria o de azar, que nos va a representar el efecto conjunto de muchas pequeñas causas que van a influir en el fenómeno y cuyos efectos por separado van a ser despreciables, pero que al sumarse pueden tener un efecto conjunto importante. Esto es debido a que el mundo real es demasiado complejo, y como, en cambio, para poder manejar el modelo, debemos procurar que sea cuanto más sencillo mejor, pues tendremos que seleccionar o poner de manifiesto las causas más importantes, incluyendo las demás en una variable llamada error. Esta variable será la que explicará que, aunque nosotros sepamos exactamente los valores tomados por las variables independientes, no podamos predecir, sin cometer algún error, cuál será el valor exacto de la variable de respuesta.

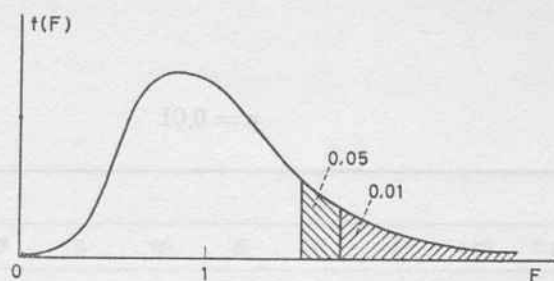
Ejemplo: Dentro de un grupo de niños, que asisten a una determinada

Distribuidor F de Snedecor

Probabilidades $\left\{ \begin{array}{l} 5\% \\ 1\% \end{array} \right.$

$$\alpha = 0,05$$

k_1	k_2											
	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
	4,052	4,999	5,403	5,625	5,764	5,859	5,928	5,981	6,022	6,056	6,082	6,106
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
	98,49	99,00	99,17	99,25	99,30	99,33	99,34	99,36	99,38	99,40	99,41	99,42
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,64	8,81	8,78	8,76	8,74
	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67



k_1												k_1
14	16	20	24	30	40	50	75	100	200	500	∞	
245	246	248	249	250	251	252	253	253	254	254	254	1
6,142	6,169	6,208	6,234	6,258	6,256	6,302	6,323	6,334	6,352	6,361	6,360	
19,42	19,43	19,44	19,45	19,46	19,47	19,47	19,48	19,49	19,49	19,50	19,50	2
99,43	99,44	99,45	99,46	99,47	99,48	99,48	99,49	99,49	99,49	99,50	99,50	
8,71	8,69	8,66	8,64	8,62	8,60	8,58	8,57	8,56	8,54	8,54	8,53	3
26,92	26,93	26,69	26,60	26,50	26,41	26,35	26,27	26,23	26,18	26,14	26,12	
5,87	5,84	5,80	5,77	5,74	5,71	5,70	5,68	5,66	5,65	5,64	5,63	4
14,24	14,15	14,02	13,93	13,83	13,74	13,69	13,61	13,57	13,52	13,48	13,46	
4,64	4,60	4,56	4,53	4,50	4,46	4,44	4,42	4,40	4,38	4,37	4,36	5
9,77	9,68	9,55	9,47	9,35	9,29	9,24	9,17	9,13	9,07	9,04	9,02	
3,96	3,92	3,87	3,84	3,81	3,77	3,75	3,72	3,71	3,69	3,68	3,67	6
7,60	7,52	7,39	7,31	7,23	7,14	7,09	7,02	6,99	6,94	6,90	6,88	
3,52	3,49	3,44	3,41	3,38	3,34	3,32	3,29	3,28	3,25	3,24	3,23	7
6,35	6,27	6,15	6,07	5,98	5,90	5,85	5,78	5,75	5,70	5,67	5,65	
3,23	3,20	3,15	3,12	3,08	3,05	3,03	3,00	2,98	2,96	2,94	2,93	8
5,56	5,48	4,36	5,28	5,20	5,11	5,06	5,00	4,96	4,91	4,88	4,86	

(Continuación)

 $\alpha = 0,01$

k_1	k_2											
	1	2	3	4	5	6	7	8	9	10	11	12
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
	9,65	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,30	4,30	4,22	4,16
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	4,00	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,65	2,59	2,55	2,51	2,48
	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38
	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34
	8,28	6,01	5,09	4,58	4,25	4,01	3,85	3,71	3,60	3,51	3,44	3,37
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38	2,34	2,31
	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,34	2,31	2,28
	8,10	5,85	4,94	4,43	4,10	3,87	3,71	3,56	3,45	3,37	3,30	3,23
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25
	8,02	5,78	4,87	4,37	4,04	3,81	3,65	3,51	3,40	3,31	3,24	3,17
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30	2,26	2,23
	7,94	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	2,32	2,28	2,24	2,20
	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,14	3,07

de la tabla anterior)

k_1	k_2											
	14	16	20	24	30	40	50	75	100	200	500	∞
9	3,02	2,98	2,93	2,90	2,86	2,82	2,80	2,77	2,76	2,73	2,72	2,71
	5,00	4,92	4,80	4,73	4,64	4,56	4,51	4,45	4,41	4,36	4,33	4,31
10	2,86	2,82	2,77	2,74	2,70	2,67	2,64	2,61	2,59	2,56	2,55	2,54
	4,60	4,52	4,41	4,33	4,25	4,17	4,12	4,05	4,01	3,96	3,93	3,91
11	2,74	2,70	2,65	2,61	2,57	2,53	2,50	2,47	2,45	2,42	2,41	2,40
	4,29	4,21	4,10	4,02	3,94	3,86	3,80	3,74	3,70	3,66	3,62	3,60
12	2,64	2,60	2,54	2,50	2,46	2,42	2,40	2,36	2,35	2,32	2,31	2,30
	4,05	3,98	3,86	3,78	3,70	3,61	3,56	3,49	3,46	3,41	3,38	3,36
13	2,55	2,51	2,46	2,42	2,38	2,34	2,32	2,28	2,26	2,24	2,22	2,21
	3,85	3,78	3,67	3,59	3,51	3,42	3,37	3,30	3,27	3,21	3,18	3,16
14	2,48	2,44	2,39	2,35	2,31	2,27	2,24	2,21	2,19	2,16	2,14	2,13
	3,70	3,62	3,51	3,43	3,34	3,26	3,21	3,14	3,11	3,06	3,02	3,00
15	2,43	2,39	2,33	2,29	2,25	2,21	2,18	2,15	2,12	2,10	2,08	2,07
	3,56	3,48	3,36	3,29	3,20	3,12	3,07	3,00	2,97	2,92	2,89	2,87
16	2,37	2,33	2,28	2,24	2,20	2,16	2,13	2,09	2,07	2,04	2,02	2,01
	3,45	3,37	3,25	3,18	3,10	3,01	2,96	2,89	2,86	2,80	2,77	2,75
17	2,33	2,29	2,23	2,19	2,15	2,11	2,08	2,04	2,02	1,99	1,97	1,96
	3,35	3,27	3,16	3,08	3,00	2,92	2,86	2,79	2,76	2,70	2,67	2,65
18	2,29	2,25	2,19	2,15	2,11	2,07	2,04	2,00	1,98	1,95	1,93	1,92
	3,27	3,19	3,07	3,00	2,91	2,83	2,78	2,71	2,68	2,62	2,59	2,57
19	2,26	2,21	2,15	2,11	2,07	2,02	2,00	1,96	1,94	1,91	1,90	1,88
	3,19	3,12	3,00	2,92	2,84	2,76	2,70	2,63	2,60	2,54	2,51	2,49
20	2,23	2,16	2,12	2,08	2,04	1,99	1,96	1,92	1,90	1,87	1,85	1,84
	3,13	3,05	2,94	2,86	2,77	2,69	2,63	2,56	2,53	2,47	2,44	2,42
21	2,20	2,15	2,09	2,05	2,00	1,96	1,93	1,89	1,87	1,84	1,82	1,81
	3,07	2,99	2,88	2,80	2,72	2,63	2,58	2,51	2,47	2,42	2,38	2,36
22	2,18	2,13	2,07	2,03	1,98	1,93	1,91	1,87	1,84	1,81	1,80	1,78
	3,02	2,94	2,83	2,75	2,67	2,58	2,53	2,46	2,42	2,37	2,33	2,31
23	2,14	2,10	2,04	2,00	1,96	1,91	1,88	1,84	1,82	1,79	1,77	1,76
	2,97	2,89	2,78	2,70	2,62	2,53	2,48	2,41	2,37	2,32	2,28	2,26

clase de una escuela, se desea explicar las notas que obtienen. Variables que se podrían hacer intervenir: cociente intelectual, número de horas de estudio el día, edad, educación de los padres, sexo, número de hermanos, etc...

Concepto de prueba de significación

Una vez que tengamos planteado el modelo matemático, y hayamos observado los valores tomados por las variables independientes, podremos predecir o dar un valor aproximado para la variable de respuesta; suponiendo que más tarde podamos observar el valor tomado por esta variable. De la comparación entre el valor real y el valor esperado, deduciremos si el modelo es apropiado o no para representar nuestro fenómeno.

La prueba de significación será una regla objetiva de comparación entre los valores reales y los valores esperados.

Esta regla objetiva resultará de la aplicación del llamado Análisis de la varianza al modelo planteado.

Empezaremos por estudiar el Análisis de la varianza para una clasificación simple de observaciones: este análisis se referirá a experimentos con un factor simple.

Experimentos con un factor simple sin restricciones en la aleatoriedad

Supongamos que nosotros queremos contrastar si un determinado factor interviene en un modelo. Lo que haremos es ir dándole a este factor distintos niveles (distintos valores) y viendo las variaciones que se producen en la variable de respuesta. Si dedujésemos que estas variaciones no son significativas (despreciables), es que el factor no nos influiría en el modelo.

Tendremos en este caso un único factor e indicaremos por T_j : $j = 1 - K$, el efecto de aplicar el factor en su nivel j . Para cada nivel del factor efectuaremos n_j observaciones.

El modelo será

$$X_{ij} = \mu + T_j + E_{ij}$$

donde X_{ij} representa la observación i -ésima de la variable de respuesta correspondiente al nivel j del factor.

Es un efecto común que aparecerá en todas las observaciones y que no depende del factor considerado.

T_j se llamará efecto del tratamiento j .

E_{ij} error aleatorio que aparece en la observación i con el tratamiento j .

Se consideran usualmente las variables E_{ij} como variables aleatorias normales o independientes de media cero y varianza la misma a todos los niveles:

$$E_{ij} : N(0, \sigma_e^2)$$

Se considera μ como parámetro o valor fijo y: $T_1, T_2 \dots T_k$ como parámetros fijos cuando los niveles de tratamiento lo son.

Se supone que $\sum_{j=1}^k T_j = 0$, esta hipótesis se hace porque con ella si contrastamos que todos los tratamientos T_j son nulos, esto equivaldrá a contrastar que son iguales entre sí (a veces lo que interesa no es contrastar que el factor no interviene, sino que no importa el nivel que le demos; que los efectos de todos los niveles son iguales).

Tal y como hemos planteado el modelo, si la hipótesis que llamaremos H_0 de que $T_j = 0$ para todo j es cierta, nos resulta que cada observación proviene de una población normal de media μ y varianza σ_e^2 .

Análisis de la varianza

Recogidos los datos se consideran K poblaciones cada una correspondiente a un tratamiento.

	Tratamiento		
observaciones	X_{11}	X_{12} X_{1k}	
	X_{21}	X_{22} X_{2k}	
	⋮	⋮	
	X_{n11}	X_{n22} X_{nkk}	
Totales	T_{o1}	T_{o2} T_{ok}	
número de observaciones	n_1	n_2 n_k	
medias muestrales	\bar{X}_{o1}	\bar{X}_{o2} \bar{X}_{ok}	

donde $T_{oj} = \sum_{i=1}^{n_j} T_{ij}$ representa la suma total de las observaciones hechas con el tratamiento j .

n_j : número de observaciones hechas con el tratamiento j .

\bar{X}_{oj} la media muestral de estas observaciones: $\bar{X}_{oj} = \frac{T_{oj}}{n_j}$.

Planteemos:

$T_{..} = \sum_{j=1}^k T_{.j}$; $T_{..}$ representa la suma total de las observaciones hechas.

$N = \sum_{j=1}^k n_j$; N representa el número total de observaciones.

$\bar{X}_{..} = (\sum_{j=1}^k n_j \bar{X}_{.j})/N$; $\bar{X}_{..}$ representa la media muestral de las observaciones.

Ahora nosotros consideramos la identidad:

$$\bar{X}_{jj} - \bar{X}_{..} = (\bar{X}_{.j} - \bar{X}_{..}) + (\bar{X}_{ij} - \bar{X}_{.j})$$

Elevando al cuadrado y sumando queda:

$$\begin{aligned} \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{..})^2 &= \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{X}_{.j} - \bar{X}_{..})^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{.j})^2 = \\ &= \sum_{j=1}^k n_j (\bar{X}_{.j} - \bar{X}_{..})^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{.j})^2 \end{aligned}$$

pues el término correspondiente a los dobles productos resulta ser nulo.

El primer miembro de esta igualdad dividido por N representa la varianza total de las observaciones; por ello lo que hemos hecho al plantear esta igualdad es una descomposición de la varianza total en dos sumandos, de los cuales, el segundo representa salvo un factor constante la media de las varianzas dentro de las K submuestras consideradas, mientras que el primero, salvo un factor constante, representa la varianza entre las medias de estas submuestras.

De la descomposición anterior de la varianza y del hecho de que si la hipótesis $H_0: T_j = 0 \quad j = 1 - K$ es cierta, X_{ij} será una variable Normal: $N(\mu, \sigma_e^2)$, resulta por aplicación del teorema de Cochran (1) sobre suma de formas cuadráticas de variables normales independientes e idénticamente distribuidas que haciendo:

$$Q = \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{..})^2$$

$$Q_1 = \sum_{j=1}^k n_j (\bar{X}_{.j} - \bar{X}_{..})^2$$

$$Q_2 = \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{.j})^2$$

las variables $\frac{Q_1}{\sigma_e^2}$, $\frac{Q_2}{\sigma_e^2}$ y $\frac{Q}{\sigma_e^2}$ son variables aleatorias independientes que

siguen distribuciones Chi-cuadrado: X^2 con $(N-1)$, $(K-1)$ y $(N-K)$ grados de libertad respectivamente.

Deduciéndose de aquí que si consideramos el cociente:

$$(1) \quad F = \frac{\sum_{j=1}^k n_j (\bar{X}_{.j} - \bar{X}_{..})^2 / (K - 1)}{\sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{.j})^2 / (N - K)}$$

Esta variable F , siempre que H_0 sea cierta, se distribuirá con una distribución F de Snedecor con $(K - 1)$, $(N - K)$ grados de libertad.

Forma de actuar

Nosotros vamos a fijar un número α , $0 \leq \alpha \leq 1$, llamado nivel de significación, que va a representar la probabilidad de rechazar la hipótesis H_0 siendo cierta, es decir, puede suceder que realmente todos los tratamientos sean iguales y, que, no obstante, debido a los errores de observación E_{ij} , dé la casualidad de que éstos sean muy grandes, que las observaciones X_{ij} sean muy dispares y que nosotros vayamos a deducir que los tratamientos no son iguales.

Este número α lo fijaremos tanto más pequeño cuanto más importante sea para nosotros no equivocarnos y no cometer el tipo de error expuesto.

Prefijado α , la distribución F de Snedecor es una distribución que está tabulada, y entonces en la tabla correspondiente $(K - 1)$, $(N - K)$ grados de libertad, podemos buscar una abscisa F_α tal, que la probabilidad de que la variable F tome un valor mayor que F_α sea igual a α .

Tenemos pues $P [F > F_\alpha] = \alpha$.

Ahora con nuestros datos calcularemos el valor tomado por F en nuestro problema según la fórmula (1).

Entonces, si F resulta mayor que F_α rechazamos la hipótesis H_0 de que todos los tratamientos sean iguales, ya que de no hacerlo sería aceptar que se nos había presentado un suceso de probabilidad tan pequeña como α , cosa que no estamos dispuestos a admitir.

La regla será pues si $F > F_\alpha$, negamos que todos los tratamientos sean iguales y decimos que los resultados son significativos, si $F < F_\alpha$ decimos que los resultados, al nivel de significación considerado, no son significativos de que la hipótesis sea falsa y la aceptamos: a la vista de los resultados es admisible que todos los tratamientos sean iguales.

La idea intuitiva del test F sería la siguiente:

Cuando disponemos de dos o más muestras aleatorias de una población normalmente distribuida, tenemos dos estimaciones independientes de la varianza poblacional, una basada en la media de las varianzas dentro de cada muestra $(Q_2/N - K)$, y otra basada en la varianza entre las medias de las muestras $(Q/K - 1)$. Podemos esperar que, generalmente, estas dos estimaciones sean casi iguales. Pero supongamos ahora que las diversas muestras

proviene de distintas poblaciones con medias diferentes. En este caso la estimación basada en la varianza dentro de las muestras es una estimación de la varianza poblacional. Pero la estimación, basada en las diferencias entre las medias, es mayor: refleja no sólo la varianza poblacional, sino también una variación adicional debida a las diferencias reales de las medias poblacionales.

Ahora, F , es sencillamente el cociente entre las dos estimaciones de la varianza.

$$F = \frac{\text{estimación de la varianza basada en las diferencias entre medias}}{\text{estimación de la varianza basada en la varianza dentro de las muestras}}$$

Si el cociente es pequeño las dos estimaciones concuerdan bastante. Luego, debemos deducir que las diferencias de medias observadas reflejan la variación muestral de una población única con una sola media. Si el cociente es grande, debemos al menos intuir que las medias observadas difieren tanto que pueden representar diferencias reales de los valores poblacionales.

Ejemplo

Como parte de un amplio estudio sobre psiquiatría preventiva, Muuss examinó la relación entre un programa de aprendizaje causal y la ansiedad manifiesta, en una muestra de estudiantes de sexto grado de enseñanza primaria.

La hipótesis básica del estudio es que los niños preparados para comprender "las interacciones, múltiples y complejas, de los factores que influyen en la conducta humana" presentarán signos de mejor salud mental que los demás. Uno de los índices de salud mental estudiados fue el grado en que un niño prefiere, en su trato con otros, ser intransigente o agresivo, en lugar de comprensivo.

Se seleccionaron tres muestras de estudiantes, homogéneas en cuanto al sexo, la edad y el cociente intelectual. Un primer grupo de 25 fue sometido a un aprendizaje causal de dos años, el segundo a uno de un año y el tercer grupo no recibió ninguna preparación. Se sometió a los 75 estudiantes a un test característico de agresividad.

Los resultados figuran en la tabla siguiente:

Niveles de agresividad de los estudiantes según el aprendizaje causal

	Aprendizaje causal			Total grupo
	A Ninguno	B Un año	C Dos años	
$\bar{X}_{.j}$	4,12	5,00	2,04	3,72
$\bar{X}_{.j} - \bar{X}_{..}$	0,40	1,28	-1,68	
$(\bar{X}_{.j} - \bar{X}_{..})^2$	0,16	1,63	2,82	
n_j	25	25	25	
$n_j(\bar{X}_{.j} - \bar{X}_{..})^2$	4,00	40,75	78,50	
$\sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{.j})^2$				607,12

Hipótesis H_0

El aprendizaje causal no influye sobre los niveles de agresividad. Si esta hipótesis es cierta, la expresión F dada por (1) se distribuirá según una distribución F de Snedecor con $K - 1 = 3 - 1 = 2$ grados de libertad y $N - K = 75 - 3 = 72$ grados de libertad.

Nivel de significación

Se fija $\alpha = 0'01$.

Región de rechazo

En la tabla de la F de Snedecor para $\alpha = 0'01$, y correspondiente a los grados de libertad 2 y 60 que son los más próximos a los que nos dan que figuran en la tabla, se lee $F_\alpha = 4'98$.

Decisión

Calculemos la expresión (1):

$$\sum_{j=1}^k n_j (\bar{X}_{.j} - \bar{X}_{..})^2 = 4'00 + 40'75 + 78'50 = 123'25$$

$$\sum_{i=1}^k \sum_{j=1}^{n_j} (X_{ij} - \bar{X}_{.j})^2 = \sum_{i=1}^k \sum_{j=1}^{n_j} (X_{ij} - \bar{X}_{..})^2 - \sum_{i=1}^k n_i (X_{.i} - \bar{X}_{..})^2 = 607'12 - 123'2 = 483'87$$

$$F = \frac{123'25}{483'87} \cdot \frac{72}{2} = 9'1$$

Como $9'1 > 4'98$, $F > F_\alpha$ y rechazamos que el aprendizaje causal no influya sobre los niveles de agresividad o lo que es igual aceptamos que influye.

Comparaciones parciales. Método de los contrastes ortogonales

Una vez sacada, como en el problema anterior, diferencia significativa entre las medidas factores, se plantea la pregunta: ¿Qué medias difieren? Puede en efecto que la tercera media difiera de la primera (dos años de aprendizaje sí influyan) y que sin embargo la segunda difiera de la primera (un año solo no influya).

Por el método anterior lo que hemos hecho es comparar todos los tratamientos a la vez, ahora vamos a dar un procedimiento para hacer comparaciones parciales entre los tratamientos.

Sean T_1, \dots, T_k los totales por tratamiento.

Se define un contraste C_m sobre el número total de tratamientos como:

$$C_m = \sum_{j=1}^k C_{jm} T_{.j}$$

siendo C_{jm} números tales que: $\sum_{j=1}^k n_j C_{jm} = 0$.

Dados dos contrastes, C_m y C_q , diremos que son ortogonales si:

$$\sum_{j=1}^k n_j C_{jm} C_{jq} = 0.$$

El método de los contrastes ortogonales consiste en elegir un número de contrastes ortogonales menor o igual que $(K-1)$ y formar para cada uno de ellos la expresión:

$$SS_{C_m} = \frac{C_m^2}{\sum_{j=1}^k n_j C_{jm}^2}$$

Entonces, si es cierto que todos los tratamientos que aparecen en la expresión de C_m son iguales entre sí, resulta:

$$F = \frac{SS_{cm}/1}{\sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{.j})^2 / (N - K)}$$

distribuida según una F de Snedecor con

1 y $(N - K)$ grados de libertad.

Ejemplo

Si nosotros, en el problema ya tratado, quisiéramos ver si existe diferencia significativa entre los alumnos que no han sido sometidos a aprendizaje y los que sí lo han desarrollado durante un año, podríamos considerar el contraste:

$$C_1 = T_{.1} - T_{.2}$$

Bastará que los tratamientos 1 y 2 sean iguales entre sí, independientemente de lo que ocurra al 3, para que:

$$F = \frac{C_1^2 / 25 \cdot 2}{\sum \sum (X_{ij} - \bar{X}_{.j})^2 / (N - K)}$$

siga una F de Snedecor con $1, (N - K)$ grados de libertad.

$$T_{.2} = 25 \cdot 5'00$$

$$C_1 = 25 \cdot 0'88$$

En nuestro caso: $T_{.1} = n_1 \bar{X}_{.1} = 25 \cdot 4'12$

$$F = \frac{(25)^2 \cdot (0'88)^2 / 25 \cdot 2}{483'87 / 72} = 1'4$$

El valor de F_α , con 1 y 72 grados de libertades para $\alpha = 0'01$, es $F_\alpha = 7'08$.

Resultando admisible la hipótesis de que un año de aprendizaje no influye sobre la agresividad.

Generalización de los resultados

Todos los resultados hasta aquí expuestos se pueden generalizar en el caso de que en lugar de un factor tengamos dos o más y vayamos combinando los niveles de los distintos factores y obteniendo observaciones para cada una de estas combinaciones.

Así, cuando tengamos dos factores independientes, tendremos la tabla de doble entrada:

		Primer factor			
		Niveles			
		1	2 m	medias
Segundo factor Niveles	1	X_{11}	X_{12} X_{1m}	\bar{X}_1
	2	X_{21}	X_{22} X_{2m}	\bar{X}_2

	K	X_{k1}	X_{k2} X_{km}	\bar{X}_k
		$\bar{X}_{.1}$	$\bar{X}_{.2}$ $\bar{X}_{.n}$	$\bar{X}_{..}$

El análisis de la varianza correspondiente es:

$$\sum_i \sum_j (X_{ij} - \bar{X}_{..})^2 = \sum_i \sum_j (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_{..})^2 + m \sum_i (\bar{X}_i - \bar{X}_{..})^2 + K \sum_j (\bar{X}_j - \bar{X}_{..})^2$$

Para ver que no existe efecto del primer factor, consideraremos la expresión:

$$F = \frac{(K-1) K \sum_j (\bar{X}_j - \bar{X}_{..})^2}{\sum_i \sum_j (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_{..})^2}$$

si la hipótesis es cierta, F se distribuirá según una F de Snedecor con $(m-1)$ y $(m-1)(K-1)$ grados de libertad.

Para ver que no existe efecto del segundo factor, consideraremos la expresión:

$$G = \frac{(m-1) m \sum_i (\bar{X}_i - \bar{X}_{..})^2}{\sum_i \sum_j (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_{..})^2}$$

que si la hipótesis es cierta, se distribuirá según una F de Snedecor con $(K-1)$ y $(m-1)(K-1)$ grados de libertad.

**PUBLICACIONES DEL I.C.E.
DE LA UNIVERSIDAD DE BARCELONA**

I. DOCUMENTOS Y TRABAJOS CICLOSTILADOS EN EL AÑO 1971

- Documento n.º 1. — M. Sigúan, Valentí, Freixa:
Algunos datos estadísticos sobre el alumnado de bachiller en Barcelona.
Septiembre 1971.
- Documento n.º 2. — Juan Estruch:
El coste familiar de la Educación. Primer Informe. Septiembre 1971.
- Documento n.º 3. — Adalberto Ferrández:
La enseñanza individualizada en la E.G.B. Primer Informe. Septiembre
1971.
- Documento n.º 4. — Jaime Sarramona:
Pedagogía Institucional. Primer Informe. Septiembre 1971.
- Documento n.º 5. — María Forns:
Pruebas para medir el conocimiento de una lengua I. Septiembre 1971.
- Documento n.º 6. — María Pla:
Pruebas para medir el conocimiento de una lengua II. Septiembre 1971.
- Documento n.º 7. — Montserrat Freixa:
El condicionamiento social de las aptitudes intelectuales y su influencia
sobre el rendimiento escolar. Primer Informe. Septiembre 1971.
- Documento n.º 8. — Fernando de Cea, Francisco Arteaga:
Informe sobre 5.º experimental en E.G.B. Septiembre 1971.
- Documento n.º 9. — Informe de las actividades del I.C.E. de la Universidad
de Barcelona, Curso 1970-1971. Septiembre 1971.

II. DOCUMENTOS Y TRABAJOS CICLOSTILADOS EN EL CURSO 1971-1972

- Documento A-10. — Pedro Batallé:
Estadísticas Educativas. Primer Informe. Diciembre 1971.

- Documento A-11. — Juan Estruch:
El coste familiar de la Educación. Segundo Informe. Diciembre 1971.
- Documento A-12. — Jaime Sarramona:
Ensayo de Pedagogía Institucional. Segundo Informe. Diciembre 1971.
- Documento A-13. — Adalberto Ferrández:
Enseñanza Individualizada. Segundo Informe. Diciembre 1971.
- Documento A-14. — Juan Mestres:
Los problemas Psicopedagógicos planteados por el Bilingüismo. Primer Informe. Diciembre 1971.
- Documento A-15. — José María Doménech:
Tabulación de encuestas: Tablas a simple y doble entrada para el análisis de los resultados. Marzo 1972.
- Documento A-16. — E. Torá, J. Lahosa, M. Porter, S. Mallas:
Seminario de Cine Infantil (Psicopedagogía de la Imagen). Marzo 1972.
- Documento A-16 bis. — José M.^a Doménech:
Tratamiento de la Información con Ordenadores digitales. Junio 1972.
- Documento A-17. — José M.^a Doménech:
Programa CORR. Cálculo de medias. Desviaciones tipo. Coeficiente de correlación lineal y coeficientes de correlación múltiple. Manual de utilización. Junio 1972.
- Documento A-18. — José M.^a Doménech:
Programa BARE. Baremación de tests psicológicos, cálculo del baremo, características estadísticas y prueba de normalidad de la distribución. Manual de utilización. Junio 1972.
- Documento A-19. — Informe de las actividades del I.C.E. de la Universidad de Barcelona. Curso 1971-1972. Septiembre 1972.
- Documento A-19 bis. — José M.^a Doménech:
Programa de análisis factorial. Programa AFAC. Manual de utilización. Junio 1972.
- Documento A-20. — Jaime Sarramona:
Ensayo de Pedagogía Institucional. III. Septiembre 1972.
- Documento A-21. — Juan Mestres:
Problemas Psicopedagógicos del Bilingüismo. II. Septiembre 1972.
- Documento A-22. — Jornadas de Psicopedagogía de la Imagen Visual. Noviembre 1972.
- Documento A-23. — Javier Barragán:
Elementos de Matemáticas para profesores de E.G.B. Septiembre 1972.

III. DOCUMENTOS CICLOSTILADOS EN EL CURSO 1972-1973

- Documento A-24. — Adalberto Ferrández:
Enseñanza Individualizada. Mayo 1973.
- Documento A-25. — Seminario sobre Educación no-directiva.
- Documento A-26. — Informe de las actividades del I.C.E. de la Universidad de Barcelona. Curso 1972-1973.
- Documento A-27. — María Forns:
Organización y funcionamiento de un servicio de orientación escolar en la E.G.B.
- Documento A-28. — Joaquín Arnau:
Problemas psicopedagógicos del bilingüismo. III.
- Documento A-29. — José María Doménech:
Descripción y valoración de un curso experimental de Métodos estadísticos para estudiantes de Psicología.
- Documento A-30. — Jorge Oliver:
Sensibilización auditiva.
- Documento A-31. — E. Torá, M. Porter y otros:
Comprensión del cine por los niños.
- Documento A-32. — E. Torá y J. M. Rueda:
Escuela y Socialización (Resumen de un Seminario Permanente para Profesores de E.G.B.).
- Documento A-33. — Informe de las actividades del I.C.E. de la Universidad de Barcelona. Curso 1973-1974.
- Documento A-34. — María de Borja Solé:
La experiencia comparativa en el aprendizaje del francés.

IV. — INFORMES PUBLICADOS EN EL CURSO 1971-1972

- Informe n.º 2. — Miguel Siguán, Álvaro Monferrer:
La atención a los deficientes mentales en la provincia de Barcelona. Noviembre 1971.
- Informe n.º 3. — M. Siguán, E. Bosch, J. Subirós:
Evolución del alumnado en la Facultad de Filosofía y Letras. 1939-1971. Noviembre 1971.
- Informe n.º 4. — Escuela Universitaria del Profesorado de E.G.B. Plan de estudios del curso experimental. Noviembre 1971.
- Informe n.º 5. — Miguel Siguán, Pedro Batallé:
Composición del alumnado en la Facultad de Filosofía y Letras. Curso 1971-1972.

- Informe n.º 6. — Enrique Torá:
Hacia una metodología de la comunicación.
- Informe n.º 7. — Fernando de Cea:
Organización y funcionamiento de un Centro Piloto de E.G.B.
- Informe n.º 8. — Pedro Batallé:
Seminario de Planificación Universitaria.
- Informe n.º 9. — Vicente Benedito, José J. Piquer:
La formación pedagógica del profesorado de Bachillerato.
- Informe n.º 10. — Metodología científica aplicada a la investigación educativa. Diciembre 1974.

En prensa:

- Informe n.º 11. — El Centro Piloto de E.G.B. Font den Fargas (2.º informe).

