

El error experimental

VARIACIÓN Y ERROR

Cuando el investigador ha concluido su experimento y reúne los resultados de sus observaciones, puede constatar fácilmente, que las medidas de sus observaciones no son iguales; existe sin duda una variación. ¿Por qué aparecen estas variaciones en las medidas que ha obtenido de su experimento? Preguntarse el porqué de estas variaciones es lo mismo que preguntarse por las causas de la variabilidad de los datos. Ello nos lleva, en definitiva, al estudio de las *fuentes de variación*.

Como ya indicamos al hablar de la varianza, los resultados de un experimento pueden estar afectados por dos causas (o fuentes) fundamentales de variación:

1. La primera y más importante, puesto que obedece a los propósitos del experimento, tiene su origen en la diversidad de tratamientos experimentales, es decir, en los diversos valores atribuidos a la variable independiente.

2. La segunda fuente de variación abarca el conjunto de factores extraños al propio experimento, y que operan en la situación total. La acción de estos factores extraños puede provocar también una variación en los resultados. La variación total debida a la presencia y acción de estos factores extraños recibe el nombre de error experimental.

En resumen, podemos afirmar que la variación que observamos en los resultados de los experimentos puede atribuirse a dos fuentes principales: la variación debida a los tratamientos o condiciones experimentales y la variación debida al error experimental. En toda prueba estadística, como hemos podido comprobar, se comparan estas dos fuentes de variación.

CAUSAS PRINCIPALES DEL ERROR EXPERIMENTAL

Puede afirmarse que son tres las causas principales determinantes del error experimental:

1. La variabilidad que presentan los sujetos al reaccionar de diversas maneras ante una misma situación experimental. Sin duda alguna, existen diferencias individuales entre los diversos sujetos experimentales y es obvio también, que no todos reaccionen de forma igual. Por tanto, una primera fuente de error está determinada por la propia índole del material experimental y a las diferencias reactivas que presentan los sujetos.

2. Un segundo determinante del error experimental procede del experimentador. Este error está determinado o bien por una falta de uniformidad en la realización del experimento, o bien porque el experimentador no ha ejercido un adecuado control en los factores extraños, etc. La acción del experimentador puede favorecer, en algunos casos, el rendimiento de un grupo porque ha presentado mejor las instrucciones o porque ha dado más tiempo para la realización de la tarea.

3. Por último, una tercera fuente de error, ocurre por las técnicas de medición y registro de resultados. Las mediciones nunca suelen ser absolutamente exactas, no sólo porque los instrumentos de medida psicológicos no suelen ser muy exactos, sino también porque se pueden cometer incorrecciones en el momento de efectuar las mediciones.

Toda esta serie de determinantes del error actúan en la variable de medida, siendo a su vez la causa de que un grupo de sujetos, sometidos a una misma condición experimental, presenten una variabilidad en sus puntajes. El hecho de que en todo experimento se presenten estas clases de errores no debe dejarnos impasibles, admitiendo como algo irremediable la existencia de tales errores, sino todo lo contrario, deberemos esforzarnos por utilizar todos los medios a nuestro alcance para lograr una máxima reducción del error experimental. Sólo así, aumentaremos el poder para probar nuestra hipótesis.

TÉCNICAS DISPONIBLES PARA LOGRAR LA REDUCCIÓN DEL ERROR EXPERIMENTAL

Antes de pasar al estudio de las técnicas que podemos utilizar para conseguir una mayor exactitud en los resultados y consecuentemente, una reducción de la varianza debida al error, analizaremos mediante un ejemplo la importancia del mismo en un tipo de prueba de inferencia como la *t* de Student.

Supóngase que se ha constatado una diferencia de medias entre dos grupos experimentales de sujetos. Analicemos las consecuencias del error para dos posibles situaciones:

- En la primera situación nos encontramos con una gran varianza debida al error; por ejemplo, 5.
- En el segundo caso, la varianza debida al error es pequeña; por ejemplo, 2.

Al calcular la razón t para estas dos situaciones hipotéticas, tendremos:

$$a) t = \frac{5}{5} = 1.0; \quad b) t = \frac{5}{2} = 2.5.$$

Ante estos supuestos resultados, deberíamos concluir que con una misma diferencia de medias, se acepta H_0 en (a), y se rechaza en (b). En todo razonamiento inferencial de capital importancia conseguir una t grande, a fin de poder demostrar la hipótesis experimental. Como es evidente, esto no quiere decir que se pretenda rechazar la hipótesis nula a toda costa. Lo único que se ha intentado demostrar con este ejemplo es que si la hipótesis nula debe ser rechazada, en la primera situación no se consigue; en la segunda sí. Todo depende, en definitiva, del tamaño del error. De ahí que nuestro propósito sea el de conseguir que la varianza debida al error se reduzca lo suficiente para que se pueda rechazar la hipótesis nula (siempre que ésta sea rechazable). Si después de haber reducido el error al máximo (hasta donde nos sea posible), no podemos rechazar la hipótesis nula, podremos concluir con mucha más confianza que en nuestro experimento la intervención de la variable independiente no ha sido efectiva.

No se trata, pues, de forzar la situación de modo que nos lleva al rechazo de H_0 sea como fuere; sólo se pretende aumentar la posibilidad de rechazarla si realmente éste es el caso.

Desde hace algunos años, la metodología experimental está estudiando algunos procedimientos para lograr una mayor exactitud en los resultados de los trabajos experimentales, al intentar reducir al máximo el error experimental. A tal respecto, Cochran y Cox (1957), nos ofrecen una serie de procedimientos que vamos a comentar brevemente.

Aumento del tamaño del experimento

Sea cual sea el origen de los errores experimentales, la repetición de un experimento tiende a disminuir, de un modo gradual, los errores asociados a la diferencia entre los resultados medios de dos tratamientos; siempre, claro está, que se hayan tenido las debidas precauciones (como

por ejemplo, la aleatorización, para asegurar que un tratamiento no haya sido más favorecido en una repetición que en otra, de tal forma que los errores que afecten a cualquier tratamiento tiendan a anularse al aumentar el número de repeticiones).

Sabemos, por teoría estadística, que los errores experimentales son medidos mediante el cálculo de la varianza del error por grupo experimental. La raíz cuadrada de la varianza del error recibe el nombre de error típico.

Supóngase que se ha llevado a cabo un experimento en el que deben compararse las medidas de dos grupos y se ha obtenido una determinada varianza de error (σ^2) por grupo experimental, realizándose del mismo r repeticiones. En este caso, el cálculo de la varianza de error de la diferencia entre las medidas de los tratamientos será el siguiente:

$$V_e = \frac{2\sigma^2}{r},$$

y el correspondiente error típico:

$$\sigma = \frac{\sqrt{2\sigma^2}}{\sqrt{r}}.$$

De ello se infiere que el error experimental asociado a la diferencia entre medias de dos tratamientos tiende a disminuir en razón inversa a la raíz cuadrada del número de repeticiones. Esto constituye una aplicación de las nociones ya conocidas de varianza del error y error típico, en una situación en que éste se reduce en función del número de repeticiones, de la misma forma que el error típico estimado a partir de una sola muestra, disminuye en función del tamaño de la misma.

Si bien esto constituye una técnica idónea, en la práctica los experimentos no suelen replicarse, debido a la falta de interés que ello suscita. Sin embargo, es muy aconsejable la réplica, no sólo desde el punto de vista de reducción del error, sino como vía indirecta de confirmación de hipótesis.

Utilización de una técnica más perfeccionada

El empleo de una buena técnica depende, básicamente, del experimentador, y con ella se alcanza una serie de objetivos básicos:

- Asegurar la uniformidad en la aplicación de los tratamientos experimentales. Así, por ejemplo, con una buena técnica se puede medir con bastante precisión el tiempo empleado en la aplicación de una prueba, en la que todos los sujetos hayan recibido las mismas instrucciones.

2. Ejercer un adecuado control sobre los factores externos de tal manera que cada condición experimental ejerza su efecto en condiciones deseables y comparables. Es difícil llegar a conclusiones claras con respecto al grado de precisión necesaria, ya que esto depende de la forma y condiciones en las que se realiza el experimento.

3. Estudiar el sistema para conseguir un tipo de medidas precisas, que no contaminen los efectos de los tratamientos. En muchas clases de experimentos el tipo de medida a utilizar es evidente; sin embargo, en otros, el empleo de un sistema satisfactorio de medida requiere muchos años de investigación, como ocurre en los estudios de psicología social.

4. Prevenir, por último, grandes errores, de los que ninguna clase de experimentación parece estar completamente libre. Así tenemos, por ejemplo, una adecuada supervisión del trabajo de los colaboradores, comprobación de los datos obtenidos en cada prueba, etc. De esta forma se descubren y rectifican errores a tiempo.

Debemos admitir que la utilización de una técnica inadecuada puede repercutir en los errores experimentales en dos maneras básicas:

a) La utilización de una técnica inadecuada puede introducir, en los resultados experimentales, fluctuaciones adiciones de naturaleza más o menos aleatoria. Estas fluctuaciones se revelarán en el momento de estimar el error para la prueba estadística. Así, si comparamos el error típico que ha obtenido un experimentador, con los de otros que han trabajado en el mismo experimento, y comprobamos que existen diferencias notables, deberemos buscar las posibles causas de tales diferencias antes de seguir adelante en el experimento.

b) Mediante la utilización de sistemas inadecuados de medida, podemos encontrarnos con estimaciones del error carentes de precisión. Contra estos posibles fallos, deberemos tener en cuenta no sólo la habilidad y cuidado en la construcción y manejo de los instrumentos de medición, sino también, el uso inteligente de la aleatorización.

Tomando medidas adicionales

En el momento de planificar un experimento hay que poner el máximo cuidado en la selección del material, a fin de obtener la mayor exactitud con un mínimo de tiempo y costo.

Muchas veces se intenta preparar el material según los objetivos de la investigación. Podemos citar, por ejemplo, la selección de muestras de poblaciones muy homogéneas. No obstante, este tipo de selección resulta a menudo difícil de conseguir, sobre todo en experimentos de tipo psicológico o sociológico. En estos casos, será de mucha utilidad, observar una serie de *medidas adicionales* o *suplementarias*, que nos

pueden servir para conocer la posible evolución o cambio de los grupos experimentales.

Si tenemos, por ejemplo, un experimento en el que se estudia el efecto que diferentes dietas ejercen en los animales, en cuanto al peso; los pesos que tenían los animales al principio del experimento constituyen una buena medida suplementaria de este tipo. Puesto que el incremento de peso de los animales, durante el experimento, se halla probablemente correlacionado con el peso inicial.

Mediante un procedimiento estadístico denominado análisis de la covarianza, se puede llegar a conocer, a partir de los resultados, el alcance de las influencias ejercidas por las variaciones en estas medidas suplementarias. Si ajustamos los pesos medios de cada tratamiento a los pesos que presentaron los animales al comenzarse el experimento, habremos eliminado una fuente de error: la producida por la diferencia de pesos iniciales. Así, los efectos de las variaciones en los pesos iniciales son eliminados, en gran parte, del error experimental, sin la necesidad de que sean iguales los pesos iniciales al planificar el experimento (Cochran y Cox, 1957).

La finalidad de este tipo de análisis consiste en la eliminación de errores experimentales que se derivan de fuentes extrañas de variación, cuyo control, mediante una técnica adecuada, sería, en muchos casos, muy costoso o físicamente imposible. Por esta razón, en muchos trabajos experimentales, la utilización del análisis de covarianza ha duplicado la eficiencia del experimento. El análisis de covarianza se reduce, en síntesis, a tener en cuenta medidas suplementarias, y aplicar los ajustes correspondientes a las medias de los tratamientos.

Utilización de un diseño experimental adecuado

Se puede, también, reducir poderosamente los errores experimentales mediante la utilización de un adecuado diseño experimental. Uno de los objetivos básicos del diseño experimental es la minimización del error; esta es la razón de la importancia de la elección de un adecuado diseño que sea sensible para detectar, con la mayor exactitud posible, los efectos producidos por los diferentes tratamientos.

La idea básica del diseño experimental es simple. Supóngase que se realizan de un mismo experimento un número r de repeticiones. Los errores experimentales de cualquier repetición sólo se originarán de las fuentes de variación que afecten a los grupos experimentales que forman una réplica. De ahí que la estimación de error en el caso de la diferencia entre medias de dos tratamientos para determinado número de repeticiones, dependerá, exclusivamente, de las variaciones que se produzcan en cada repetición. Consecuentemente, las variaciones que puedan

darse entre una réplica y otra no afectarán el error experimental de los resultados. Al llevar a cabo un experimento deberá tenerse en cuenta este principio básico: *el error sólo afectará los resultados dentro de los grupos experimentales de cada réplica y no a las diferencias entre las repeticiones*. Una de las consecuencias que se desprenden de este principio es que en todo experimento deberemos controlar las posibles fuentes de variación que afecten los diversos grupos experimentales que forman una réplica, sin preocuparnos de las diferencias que se produzcan entre una repetición y otra. Si, por ejemplo, tenemos una cantidad heterogénea de sujetos y los agrupamos de tal forma que los diversos grupos que constituyen una réplica sean lo más homogéneos posibles, el error experimental quedará afectado por la variabilidad existente dentro de cada réplica, no por la diferencia existente entre las réplicas. Gracias a este sistema, podremos lograr experimentos muy precisos a partir de sujetos que a primera vista parecían poco idóneos.

De igual modo, si una técnica o procedimiento experimental no puede ser mantenida a lo largo de la investigación, lo importante es conservarla uniforme para todos los grupos que constituyen una réplica. En todo caso, podrán introducirse cambios en la técnica cuando se pase de una repetición a otra, pero manteniendo su uniformidad para todos los grupos que constituyen cada réplica.

El concepto fundamental de la utilización del diseño experimental lo podemos resumir de la siguiente manera: se intentará reducir al máximo las posibles fuentes de error experimental mediante el empleo de un adecuado diseño de experimentación.

Los diseños, en relación con el control del error experimental, utilizan una serie de procedimientos que van desde la aleatorización, pasando por las técnicas de bloqueo, hasta el empleo de sujeto único, para lograr un máximo de la varianza debida al error. Dejamos para más adelante, el análisis concreto de estas técnicas, así como el estudio de los principales diseños experimentales.

EL TÉRMINO DE ERROR

Al intentar probar una hipótesis experimental hemos dicho que, estratégicamente, se parte de la hipótesis nula. En ella se supone que las diferencias existentes entre las medidas de dos grupos de igual tamaño, elegidos al azar de una misma población, son resultado de la propia selección al azar. En este caso la aleatorización ofrece al experimentador la garantía de que las leyes del azar han sido efectivas. El azar asegura que cada sujeto posea la misma probabilidad de pertenecer a cualquier grupo experimental.

Puesto que un grupo de sujetos elegidos al azar de una determinada población constituye una muestra representativa de aquella, se puede obtener una estimación de la varianza de la población a partir o bien de las diferencias existentes entre los sujetos de cada grupo (varianza intragrupo), o bien, de la diferencia entre las medidas de los grupos (varianza intergrupos). La lógica básica del razonamiento de la hipótesis nula es el siguiente: cualquier diferencia entre las medias de los grupos, o entre los sujetos de cada grupo se debe a la propia selección al azar.

Sin embargo, las medidas que se obtienen de los grupos han sido sometidas a los diferentes tratamientos experimentales. En efecto, para cualquier clase de diseño, cada grupo ha sido sometido a determinado valor de la variable independiente, esperándose, bajo la hipótesis nula, que el efecto de las distintas condiciones experimentales no diversifique, sistemáticamente, las medidas tomadas de las respuestas de los individuos. En caso de que la hipótesis nula fuera falsa, se esperaría que los diversos valores de la variable independiente afectaran de modo distinto a las conductas de los individuos y que esta acción se reflejara en la diferencia de medias entre los grupos. Sin embargo, aunque la hipótesis nula sea falsa y se comprueben diferencias significativas entre las medias de los grupos, la variación de las medidas tomadas de los sujetos dentro de cada grupo, seguirá variando como consecuencia de la aleatorización. De esta forma, la varianza intragrupo, la existente entre las medidas tomadas de los sujetos dentro de cada grupo, a pesar de la aplicación de los tratamientos, seguirá reflejando la varianza de la población. La variabilidad entre los sujetos seguirá siendo la misma, dado que todos los sujetos pertenecientes a determinado grupo han sido sometidos a idénticas condiciones experimentales. Por ello, la varianza intragrupo constituye una buena estimación de la varianza de la población.

Desde el punto de vista estadístico, la prueba de la hipótesis nula se efectúa al comparar la varianza de la población estimada a partir de la variabilidad de los puntajes dentro de cada grupo (dicha estimación no refleja el efecto de la variable independiente; sólo refleja la variación debida al azar), con la diferencia observada entre las medias de los grupos (dicha diferencia recoge la acción de los tratamientos experimentales). De hecho, las tablas de valores de las razones t y F reflejan la frecuencia con la que las diferencias en la estimación de la varianza de dos poblaciones puede esperarse ocurra al azar, para determinado número de grados de libertad, es decir, estas tablas representan una completa descripción de la hipótesis nula. Cuanto más discrepantes sean las estimaciones de las varianzas de dos poblaciones, menos probabilidad hay de que la hipótesis nula sea válida y mayor probabilidad de que la variable independiente haya ejercido un efecto en la conducta de los individuos.

ESTIMACIÓN Y CÁLCULO DEL ERROR

La varianza intragrupo es considerada como una buena estimación de la varianza de la población, puesto que no es afectada por la acción de los tratamientos. Esta varianza suele recibir el nombre de *término de error*, y sirve de criterio de contrastación de la variabilidad de las medias de los grupos. Desde el punto de vista matemático, sabemos que la varianza de una distribución de medias muestrales (procedentes de una misma población) es igual a la varianza de la población partida por el número de muestras. Esto nos ayudará a comprender la lógica de las pruebas de significación de la diferencia entre medias muestrales. En estas pruebas, ya se trate de la *t* o la *F*, se compara la diferencia entre las medias de dos grupos o tratamiento, con el error típico de dicha diferencia.

Tenemos, por tanto, que el *error típico*, como medida de las fluctuaciones al azar, es aquel término con el que se comparan los resultados de las manipulaciones experimentales. Por tanto, al utilizar, por ejemplo, una prueba de significación *t*, nos podemos preguntar si la diferencia que observamos entre las medidas de dos grupos experimentales es real, o solamente, una consecuencia del muestreo al azar. Para dar una respuesta a esta pregunta deberemos calcular el error típico de la diferencia, utilizándolo como término de comparación.

Para calcular el error típico, dividiremos la estimación de la varianza de la población entre el número de casos o muestras y, luego, hallaremos la raíz cuadrada de este valor. El error típico es, pues, una medida de la dispersión de la distribución de las medias muestrales.

Supóngase que hemos seleccionado cinco muestras, con 100 puntajes cada una, cuyas medias son, respectivamente (véase fig. 17.1):

$$\bar{X}_1 = 110$$

$$\bar{X}_2 = 109$$

$$\bar{X}_3 = 112 \quad \bar{X}_4 = 110$$

$$\bar{X}_4 = 111$$

$$\bar{X}_5 = 108$$

Si calculamos el error típico de esta distribución muestral, tendremos:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum x^2}{n}} = \sqrt{\frac{10}{5}} = \sqrt{2} = 1.414.$$

El valor obtenido, 1.414, es una estimación del error típico de la media general de la distribución. Pero si partimos del caso más frecuente

de tener una sola muestra, cuya media puede ser por ejemplo, 110 y la desviación típica, 10; también se puede estimar el error típico de la distribución muestral de la siguiente forma:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{100}} = 1.$$

Como es obvio, este valor, 1, constituye una estimación del error típico de la media general. Existe, por tanto, una pequeña diferencia

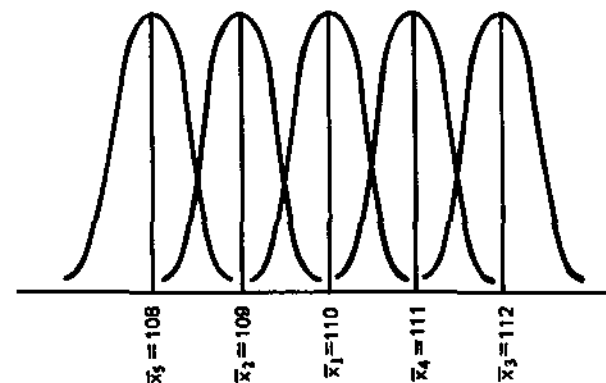


Fig. 17.1. Representación gráfica de las medias.

(0.414). Si aplicamos un tipo de prueba de significación *t*, nos encontramos que para la estimación del error típico de la diferencia de medias, intervienen dos muestras. Puesto que ambas proceden de una misma población, tomaremos una estimación conjunta de la varianza de la población, a partir de las dos muestras, para el cálculo del error típico.

Analicemos el siguiente ejemplo. Supóngase que se han elegido al azar dos grupos de niños pertenecientes a un mismo nivel escolar. Estamos interesados en estudiar la actitud de estos niños con respecto al compañerismo, después de haber pasado una película cinematográfica al primer grupo y ninguna al segundo. La media de puntajes del grupo A (grupo que vio la película) es 110; la del grupo B (no lo vio) es 100. Nos interesa saber si la diferencia entre ambos puntajes (10) es real y, por tanto, estadísticamente significativa o, por el contrario, es resultado del azar.

Para valorar el verdadero alcance de esta diferencia, calcularemos el error típico de una distribución formada por diferencias de medias. Es como si tomáramos cada posible diferencia, y fuese considerada como

un valor de X de una nueva distribución: la distribución de la diferencia de medias.

Para el cálculo del error típico de esta distribución, consideremos la estimación conjunta del error típico para cada una de las medias de los grupos.

$$\sigma_{\bar{X}_a - \bar{X}_b} = \sqrt{\sigma_{\bar{X}_a}^2 + \sigma_{\bar{X}_b}^2}$$

Supóngase que se ha realizado la experiencia cinco veces y que las diferencias de medias en cada una de ellas hubiesen sido: 10, 11, 12, 8 y 9. La media general de esta población de diferencia de medias es 10; la desviación típica, 1.414. Este valor, que es el error típico de una distribución muestral de diferencia de medias, tiene el mismo significado que el error típico de la media general para una distribución muestral de medias.

Pasemos a continuación al cálculo del error típico de la diferencia de medias a partir de los datos del ejemplo que nos ocupa. Supóngase que las desviaciones típicas obtenidas en cada grupo de niños han sido las siguientes:

$$\sigma_a = 8 \quad \text{y} \quad \sigma_b = 9, \text{ respectivamente.}$$

A continuación calculemos el error típico de las medias:

$$\sigma_{\bar{X}_a} = \frac{\sigma_a}{\sqrt{n_a}} = \frac{8}{\sqrt{100}} = 0.8,$$

$$\sigma_{\bar{X}_b} = \frac{\sigma_b}{\sqrt{n_b}} = \frac{9}{\sqrt{100}} = 0.9.$$

El error típico de la diferencia de medias será:

$$\sigma_{\bar{X}_a - \bar{X}_b} = \sqrt{\sigma_{\bar{X}_a}^2 + \sigma_{\bar{X}_b}^2} = \sqrt{0.8^2 + 0.9^2} = 1.20.$$

Si los puntajes de ambos grupos se hubiesen tomado de una tabla de números aleatorios y, por consiguiente estuviesen exentos de los tratamientos experimentales, no esperaríamos encontrar un valor significativo a esta diferencia de medias. Una diferencia como el tamaño de nuestro experimento sería un efecto del azar. El error típico de la diferencia entre medias es, por tanto, una estimación de la dispersión de tales diferencias.

Tenemos, pues, un error típico de diferencia entre medias igual a 1.20, y una diferencia igual a 10. Esto quiere decir que si hubiésemos repetido la experiencia con otras muestras, las diferencias entre los sucesivos pares de grupos, hubieran podido ser 9.80, 11.20, etc. La diferencia raramente hubiese excedido a 10 ± 1.20 . Así pues, un error típico (cuyo

valor en este caso es 1.20) indica los límites más allá de los cuales es poco probable que ocurran valores de diferencia superiores o inferiores.

A continuación analizaremos cómo se interpretan dichos resultados. El error típico, 1.20, es una estimación de las fluctuaciones del azar, y 10, es el tamaño de la diferencia entre medias obtenido en nuestra experiencia. La pregunta que se debe formular es la siguiente: ¿Puede dicha diferencia haber ocurrido al azar? Para ello comparamos la diferencia obtenida con la estimación que tenemos del azar. De este modo convertiremos esta diferencia en valores típicos.

$$t = \frac{\bar{X}_a - \bar{X}_b}{\sigma_{\bar{X}_a - \bar{X}_b}} = \frac{110 - 100}{1.20} = \frac{10}{1.20} = 8.33.$$

Esto significa que la diferencia constatada ($\bar{X}_a - \bar{X}_b$) se aleja 8.33 veces de la desviación típica para una población cuya media hipotética sea igual a cero. Teóricamente, si los niños, distribuidos aleatoriamente entre ambos grupos, no hubiesen sido diferenciados mediante una manipulación experimental, no hubiéramos obtenido diferencia alguna. En cambio, en nuestro ejemplo, se ha obtenido un valor muy elevado, si lo comparamos con 1.20, que es una estimación de las fluctuaciones debi-

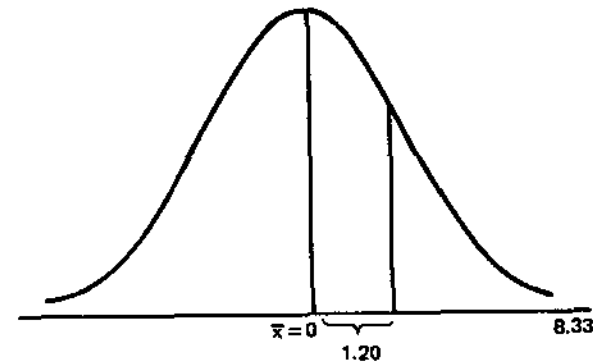


Fig. 17.2. Distribución de diferencias de medias.

das al azar. Decididamente, debemos admitir que ha ocurrido algo que escapa al puro azar y este algo es, precisamente, lo que nos interesa. Presumiblemente, el efecto de la condición experimental (visión de la película) es la causa de que haya ocurrido dicha diferencia, siempre, claro está, que se hayan controlado adecuadamente los posibles factores distorsionantes.

En la figura 17.2 tenemos una representación de la población de diferencia entre medias con media igual a 0 y desviación típica igual

a 1.20 (se establece que la media es 0 porque se supone que la media general de todas las diferencias entre medias es igual a 0).

¿Dónde se situará la diferencia de medias hallada (10) en esta gráfica? Para ello tendremos que transformarla en unidades típicas. Esto se consigue, como acabamos de ver, dividiéndola entre el error típico (1.20). Esto es lo que hemos hecho al aplicar la prueba de significación t . Mediante el cociente t convertimos una diferencia de medias en unidades típicas. Si se pretende situar el valor transformado de la diferencia de medias que hemos obtenido, fácilmente se puede comprobar que, con un nivel muy alto de probabilidad, no pertenece a la población en cuestión. Naturalmente esta diferencia podría pertenecer a dicha población, aunque esta ocurrencia es poco probable.

La misma lógica se puede aplicar en caso de que se use la prueba F , en la que también se considera el denominador de la razón como el apropiado término del error. Si los diversos valores de la variable independiente han incidido positivamente en los resultados de los respectivos grupos, es de esperar que sus medias sean considerables como procedentes de poblaciones diferentes. De ahí que, las medidas diferirían entre sí más de lo que cabría esperarse si procediesen de muestras seleccionadas al azar de una misma población. Dado que la estimación de la varianza de la población, representada por la denominación de la razón F , no queda afectada por la posible acción de los tratamientos experimentales, se puede considerar como un adecuado término del error.

El denominador de la razón F , considerado como una buena estimación de la varianza de la población, consiste en una combinación de la variabilidad producida dentro de cada grupo experimental; variabilidad, por otra parte, que no se halla afectada por la acción de la variable independiente. Por esta razón se le considera como término de error. Dentro de cada grupo experimental todos los sujetos se hallan sometidos a una misma condición, de ahí que la única variabilidad constatada se deba al muestreo al azar. Aunque las medias de los grupos varíen como consecuencia de la presencia de los diversos tratamientos, la variación intragrupo tiene como única causa la propia variación de la población, sea o no procedente la hipótesis nula. Esta varianza no afectada por las condiciones experimentales debe ser considerada, como se ha dicho, como el adecuado término de error, ya que recoge las fluctuaciones del azar que se produce en las medidas tomadas del fenómeno estudiado.

En la aplicación de cualquier prueba de hipótesis se da siempre un mismo razonamiento fundamental: el término de error, con el que se comparan las variaciones debidas a los tratamientos, se estima a partir de los elementos en que se refleja la acción del muestreo aleatorio.

RESUMEN

En este tema se ha tratado el problema de la variación del error. Se ha afirmado que cualquier variación no imputable a la acción de los tratamientos experimentales puede considerarse como error experimental: Hemos establecido tres grandes fuentes del error experimental: *a)* la variabilidad de los sujetos, que reaccionan de diferente modo ante una misma situación de pruebas; *b)* la actuación del experimentador, y *c)* las posibles técnicas de medición y registro.

A continuación se han propuesto posibles soluciones o procedimientos para la reducción del error experimental. Estos procedimientos pueden clasificarse en: *a)* aumento de las repeticiones del mismo experimento; *b)* utilización de técnicas de medida y registro mucho más precisas; *c)* recurriendo a las medidas adicionales, y *c)* la selección de un adecuado diseño de experimentación.

Por último, se ha presentado el término de error, como elemento de contrastación en las pruebas de significación estadística; asimismo, se ha visto la forma de calcularlo, su significación y su importancia en pruebas como la t de Student y la F de Snedecor.