

## Las fuentes de error en la investigación médica

ULISES ALBERTO QUESTA

Asesor en Bioestadística y Diseño  
Experimental de la SAC

En los últimos cincuenta años los conocimientos médicos se han expandido en forma realmente revolucionaria. En ello gran mérito ha tenido la investigación, apoyada en el refinamiento tecnológico que ofrece al científico una aparatología cada vez más precisa y confiable. Simultáneamente han progresado también las distintas técnicas matemáticas y de diseño experimental, capaces de determinar con exactitud el grado de confianza que merecen las conclusiones asertivas obtenidas.

La investigación médica es la investigación biológica aplicada a los problemas de la salud humana y sus objetivos pueden definirse con precisión: tratar de identificar los procesos patológicos del hombre, determinar sus causas, descubrir métodos para un mejor diagnóstico y lograr así su prevención y cura.<sup>1</sup>

La investigación médica comprende la investigación básica, que busca la elucidación de los fenómenos biológicos, sin interesarle la aplicación práctica de sus descubrimientos. Inmediatamente después se encuentra la investigación clínica, que estudia las enfermedades que afectan al hombre, sus funciones fisiológicas y los cambios patológicos observables. Más allá se sitúa la investigación médico-social, que complementa a la clínica con estudios estadísticos y demográficos que clarifican el origen y causa de muchas enfermedades. Por último, tenemos la investigación sanitaria, que estudia los mejores métodos de prevención de la enfermedad, como asimismo la mejor organización y administración de la atención médica.

Es objetivo de este trabajo ayudar al investigador médico al discutir conceptos fundamentales sobre estadística y diseño experimental, capaces de aumentar su eficiencia al proporcionarle una mayor comprensión y mejor manejo de la investigación que realiza.

### EL ERROR EXPERIMENTAL

Es un hecho habitual y forzoso que los efectos obtenidos por la aplicación de los distintos estímulos experimentales sobre los sujetos, cuando son repetidos, varíen de un ensayo a otro.\*

Por esta razón, en investigación, el problema de extraer conclu-

\* En la terminología estadística se utiliza el término "tratamiento experimental" para referirse a cualquier tipo de estímulo, lo que significa que no es sinónimo de "tratamiento terapéutico"; a su vez, "tratamiento control" no es lo mismo que "no tratamiento" (placebo) sino que también puede ser "tratamiento distinto" (droga de referencia). Los sujetos sometidos a la experiencia reciben la denominación de "unidades experimentales".

Recibido para su publicación: 3/1984  
Aceptado: 4/1984

siones de resultados no siempre coincidentes presenta dificultades que son solucionadas por la teoría estadística y de prueba de hipótesis que ofrecen aseveraciones definidas, que se caracterizan por tener una probabilidad conocida y controlable de ser correctas.

Un experimento cualquiera, sobre todo si se realiza en el campo biológico, sufre las consecuencias de la interacción de numerosos factores distintos de aquellos que son motivo de investigación; algunos de estos factores "parásitos" son total o parcialmente conocidos y, por lo tanto, mantenidos bajo el control del investigador, pero muchos otros son indiscernibles —y aun impensables— y fuente de posibles errores en la interpretación de los resultados.<sup>2</sup>

El término *error experimental* se aplica a esas variaciones espontáneas, debiendo aclararse que el término "error" no significa "equivocación", sino que involucra a todo tipo de variación no controlada.

Se acepta que existen dos fuentes fundamentales de errores experimentales:

a) La propia variabilidad espontánea del material experimental al que son aplicados los tratamientos. Es la natural variabilidad biológica que hace que cada ser vivo reaccione con características individuales al ser sometido a cualquier estímulo.

b) La falta de uniformidad en la conducción física del experimento, o sea, la incorrecta estandarización de la técnica experimental utilizada.<sup>3</sup>

Para que la existencia segura de estos errores experimentales no perturbe la tarea del investigador es necesario que éste los prevea y los mantenga bajo control, cosa que obtiene mediante un diseño adecuado del experimento y utilizando técnicas estadísticas capaces de cuantificar el monto del error deslizado, técnicas todas ellas derivadas del análisis de la varianza de los datos.

Han sido descriptos varios métodos para incrementar la exactitud de los experimentos,<sup>3</sup> los que pueden clasificarse en tres tipos:

- a) Aumentar el número de unidades experimentales o de tratamientos experimentales.
- b) Refinar la técnica experimental.
- c) Ordenar el material experimental de

manera tal que se reduzcan los efectos perjudiciales de la variabilidad espontánea.

*Número de repeticiones.* Naturalmente que si se aumenta el número de individuos sometidos a una experiencia, cualquiera sea la fuente que origine errores, la repetición disminuirá constantemente el error asociado a la diferencia entre los resultados de dos tratamientos.

Estadísticamente, el parámetro básico para medir los errores experimentales es la raíz cuadrada del cociente de la varianza del error y el número de individuos tratados (varianza del error por unidad experimental).

Asimismo, y aunque parezca difícil de creer, Fisher ha demostrado que se puede lograr un notable incremento en la exactitud si se utilizan diferentes tipos de tratamientos en el mismo experimento, en lugar de recurrir a experimentos particulares para cada tipo. Por ejemplo, si es nuestro objetivo estudiar la acción de un diurético y un hipotensor, se debe diseñar un experimento unitario donde se comparen el diurético solo, el hipotensor solo, ambos en conjunto y la ausencia de ambos, organizando un diseño factorial.

*Refinamiento de las técnicas.* Es muchas veces olvidada la importancia de lograr una adecuada exactitud y precisión en la recolección de los datos experimentales. El término *exactitud* se refiere a la seguridad con la cual la medición se aproxima al valor verdadero, en tanto que *precisión* se refiere simplemente a la capacidad de repetir las mismas mediciones.

También se debe tener en cuenta que hay que uniformar adecuadamente la aplicación de los tratamientos, controlar las influencias externas, poner todas las técnicas bien a punto, controlar firmemente el trabajo de los colaboradores y examinar personalmente los datos obtenidos en cada unidad experimental, todo esto con el fin de evitar la introducción de sesgo en la información obtenida.

*Ordenación del material experimental.* Se puede lograr una reducción marcada del error experimental ordenando las unidades experimentales de acuerdo con un eficiente diseño. Como dice Kempthorne: "El procedimiento general en la investigación científica es el formular hipótesis y ver luego si éstas se verifican directa-



mente o por sus consecuencias. Esta verificación necesita de la recolección de observaciones, y el diseño de los experimentos es esencialmente el modo en que deben ser coleccionadas esas observaciones. Diremos, brevemente, que el propósito de la teoría del diseño experimental es asegurar que el investigador obtenga los datos relevantes para su teoría de la manera más económica posible en tiempo, trabajo, dinero, etcétera".<sup>4</sup> Y, agregamos nosotros, con mayor potencia estadística y una más eficiente utilización de las unidades experimentales, *lo que permite lograr los mismos resultados con un menor número de individuos sometidos al experimento.*

Exclusión hecha de los experimentos efectuados en diseños simples al azar, que son aquellos clásicos donde a un grupo más o menos homogéneo se le opone otro de características similares y que puede ser analizado por cualquier técnica estadística, todos los demás que serán aquí discutidos necesitan invariablemente el uso de técnicas especiales basadas en el análisis de la varianza, ideado por R. A. Fisher en la década del '30.

Pero, ¿qué es la varianza? En un conjunto cualquiera de datos obtenidos durante la ejecución de un experimento los valores individuales se alejarán en magnitud variable de la media aritmética de todo el grupo, algunos superarán ese valor, otros serán inferiores a él; pues bien, ese conjunto de variaciones, esa dispersión de los datos alrededor de la media aritmética, constituye la varianza. En términos más técnicos diríamos que es el cociente del cuadrado de la suma de todas esas variaciones individuales y el número de efectivos del experimento (menos uno). La raíz cuadrada de esa cifra es el conocido desvío standard, o desviación típica.

Cuando hay dos o más grupos analizados concomitantemente, cualquier técnica estadística (diferencia entre medias, prueba t de Student, etcétera) es capaz de determinar las variaciones *entre los grupos* experimentales, pero la técnica creada por Fisher analiza también las variaciones *dentro* de cada grupo en particular (también llamada "varianza del error"); la suma de ambas será la varianza total.

Los errores experimentales de los resultados

de cualquier repetición nacen, solamente, de variaciones originadas dentro de ese grupo particular de repeticiones, no de las observadas entre los distintos grupos.

Al llevar a cabo un experimento se utiliza ese simple hecho y se trata entonces de controlar las fuentes de variación que afecten a las unidades experimentales del mismo grupo.<sup>3</sup> Por ejemplo, si tratamos hipertensos arteriales experimentalmente no intentaremos homogeneizar los grupos seleccionando pacientes de tensiones arteriales similares en cierto grado adoptando criterios de inclusión rígidos, sino que crearemos "bloques" de pacientes con un determinado grado de hipertensión (leves, medianos, severos) a los que aplicaremos el o los tratamientos bajo experimentación. De esa manera "controlamos" la heterogeneidad intergrupo al aumentar la homogeneidad intragrupo, lo que proporciona ganancia en potencia estadística y en la generalización de las conclusiones. Este tipo de diseño, conocido con el nombre de *bloques al azar*, es particularmente útil en la investigación clínica.

Existen otras numerosas maneras de distribuir el material experimental organizando diseños muy variados, adaptados todos a los cambiantes problemas que se le pueden presentar a un investigador; algunos de ellos se mencionan en el Cuadro 1, donde igualmente figuran varias técnicas estadísticas especializadas muy útiles, pero lamentablemente poco utilizadas.

Para terminar exponemos un experimento simulado con el fin de ejemplificar todo lo que hemos dicho. Supongamos que hemos recogido las cifras de la tensión arterial sistólica de tres grupos de pacientes diferentes (A, B y C), con seis repeticiones o individuos en cada uno de ellos, y deseamos saber si las diferencias observadas son estadísticamente significativas.

Luego de resistir la tentación de efectuar pruebas pareadas (prueba t) entre grupos (A versus B, B versus C y A versus C) llegamos a la conclusión de que es más razonable efectuar una sola prueba que analice simultáneamente la totalidad de los datos: el análisis de la varianza. El diseño utilizado en el experimento de marras es el denominado *simple al azar*.

Si el lector efectuara los cálculos matemáticos

**Cuadro 1**  
Diseños experimentales más habitualmente utilizados  
en la investigación médica

<i>Características del material experimental</i>	<i>Diseño más conveniente</i>
Grado de homogeneidad del material experimental	Alto - Totalmente al azar Bajo - Bloques al azar
Importancia del orden de aplicación del estímulo experimental	Secuencia con intercambio Cuadro latino
Actúan concomitantemente dos estímulos experimentales sobre el sujeto o unidad experimental	Cuadro latino Diseño factorial
Pueden variar significativamente ambos miembros de la ecuación experimental (unidades o estímulos)	Diseño anidado
Existe algún grado de correlación entre las variables	Correlación y regresión Regresión múltiple Covarianza
Existe algún tipo de interacción entre las variables o estímulos experimentales	Diseño factorial
Se analizan simultáneamente dos o más variables, en dos o más grupos, para determinar si los grupos difieren	Análisis multivariado
Se desea determinar una escala taxonómica.	Análisis multivariado con función discriminante

Cada diseño tiene sus exigencias propias, las que deben ser conocidas por el investigador para así lograr la máxima potencia y eficiencia.

necesarios observará que la conclusión final es que no existen diferencias significativas entre los grupos (Tabla 1).

Pero si se percatara que entre los seis individuos de cada grupo los tres primeros presentan valores sistemáticamente más bajos que los tres últimos y reanalizara los datos, ya organizados en un diseño en *bloques al azar*, encontrará, quizá sorprendido, que sí existen diferencias significativas ( $P < 0,05$ ) entre los grupos: las tensiones arteriales sistólicas del grupo A son

**Tabla 1**  
Datos del experimento simulado en diseño simple al azar  
y en bloques al azar

	<i>Grupo A</i>	<i>Grupo B</i>	<i>Grupo C</i>	
	115	100	101	
	113	110	108	
	116	98	102	
	145	125	130	
	150	145	142	
	160	142	158	
$\bar{x}$	133,2	120,0	123,5	
P: No significativo entre grupos A, B y C				
<i>Bloques</i>				
	<i>Grupo A</i>	<i>Grupo B</i>	<i>Grupo C</i>	$\bar{x}$
Jóvenes	115	100	101	
	113	110	108	107,0
	116	98	102	
Adultos	145	125	130	
	150	145	142	144,1
	160	142	158	
$P < 0,05$ entre grupos A, B y C				

significativamente superiores a las de los grupos restantes.

Ha ocurrido que el nuevo diseño, organizando los datos de cada grupo en dos subgrupos homogéneos, ha logrado disminuir drásticamente el grado de error experimental, muy alto en el primer diseño.

Para lograr el mismo resultado con el diseño inicial necesitaríamos aumentar el número de pacientes de... ¡18 a 90! Surge de inmediato el enorme ahorro de tiempo y esfuerzo que ello significa.

Pero hay algo más todavía. El coeficiente de variación en el diseño en bloques al azar (6,3%) es sensiblemente menor que en el diseño simple (17,3%), lo que torna más confiable el resultado



obtenido con el primero.

Técnicamente decimos que para el mismo número de tratamientos y repeticiones, el número de grados de libertad en la estimación del error cambia con el diseño, o dicho en otras palabras: *variando el diseño aumenta o disminuye la sensibilidad del experimento.*

Los criterios de agrupamiento son variados y deben responder a las necesidades particulares de cada experimento. Algunos de los más comunes son la edad, el sexo, el peso o la constitución genética; también el mismo paciente puede constituir un bloque si los tratamientos le son aplicados en forma sucesiva.

Cuando se afirma erróneamente que la hipótesis nula ( $H_0$ ) es falsa (o sea, se acepta una inexistente diferencia significativa entre las muestras) se comete el llamado error tipo I o tipo  $\alpha$ ; en el caso opuesto, al aceptarse que la hipótesis nula es cierta, siendo en realidad falsa (o sea, se rechaza una real diferencia significativa), se comete el error tipo II o tipo  $\beta$ .

La probabilidad de cometer el error tipo I está limitada por el nivel de significación aceptado (habitualmente  $P < 0,05$  ó  $0,01$ ); la de cometer el error tipo II queda supeditada al tamaño de la muestra ( $n$ ).

Debe quedar en claro que en cualquier inferencia estadística se puede cometer uno u otro tipo de error, debiendo el investigador mantener un delicado equilibrio entre ambas probabilidades.<sup>5,6</sup>

Como lo hemos demostrado en el ejemplo numérico, el riesgo de cometer el error tipo II al efectuar experimentos con una muestra demasiado pequeña es real, y se lo debe tener siempre presente, pero ante la imposibilidad de

aumentar el número de efectivos —hecho frecuente en medicina— un adecuado diseño experimental hará mucho por superar el problema.

Cochran y Cox llegan a afirmar: “Los incrementos en eficiencia, posibles de obtener por medio de la hábil construcción de los bloques, cuando son medidos en términos de repeticiones efectivas aumentadas, son frecuentemente grandes; se pueden citar ejemplos en los que la repetición efectiva ha sido aumentada de dos a diez veces. Frecuentemente, aparte del cuidado y atención dedicados preliminarmente, un diseño exacto usa el mismo material y requiere las mismas operaciones físicas que el diseño al cual sustituye”.<sup>3</sup>

Se puede afirmar que, con un inteligente uso de la estadística y del diseño experimental, el investigador incrementa en un 30% a 40% su eficiencia como tal, sin aumentar correlativamente la complejidad de sus tareas; muy al contrario, manteniendo así un mejor control de los datos obtenidos y de la investigación toda.

#### BIBLIOGRAFIA

1. Karpas J: The co-ordination of patient care and clinical investigation. *World Hospital* 5: 212, 1969.
2. Mather K: *Análisis Estadístico en Biología*. Ed Paraninfo, Madrid, 1971.
3. Cochran WG, Cox GM: *Experimental Designs*, 2ª ed. John Wiley & Sons, Inc, New York, 1962.
4. Kempthorne O: *Design and Analysis of Experiments*, 2ª ed. John Wiley & Sons, Inc. London, 1960.
5. Siegel S: *Estadística No Paramétrica*, 2ª ed. Editorial Trillas SA. México, 1979.
6. Freiman JA, Chalmers TC, Smith H, Kuebler RR: The importance of Beta, the type II error and sample size in the design and interpretation of the randomized control trial. *New Engl J Med* 299: 690, 1978.