

## Importancia del poder estadístico y el tamaño del efecto

André Chocó Cedillos

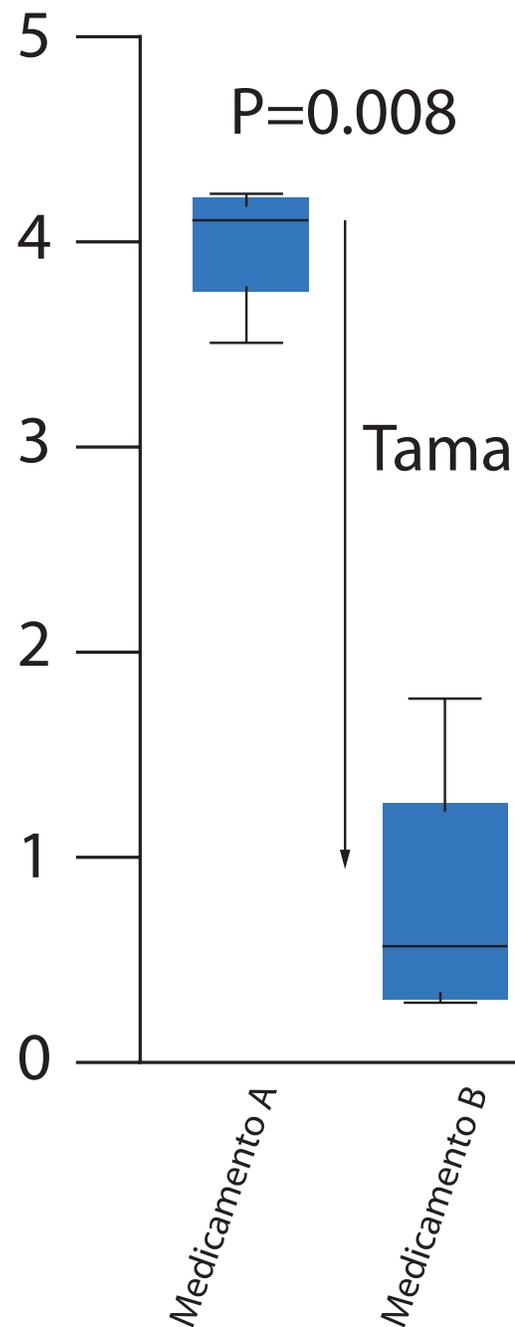
Químico Farmacéutico y asesor estadístico

panteisme@yahoo.com

La investigación en ciencias de la salud tradicionalmente ha basado sus procedimientos de análisis de datos en las pruebas de hipótesis que generan valores de probabilidad para evaluar si alguna o algunas variables de interés cambian significativamente, según los grupos de comparación o categorías de variables independientes, es decir que ciertas variables tienen un efecto sobre otras, y que ese efecto tiende a manifestarse de forma similar al replicar un experimento o realizar observaciones en condiciones similares. Por ejemplo, se realiza la comparación de dos intervenciones farmacológicas: el medicamento "A" y el medicamento "B", se desea determinar si ambos medicamentos proporcionan el mismo alivio medido en porcentaje de pacientes recuperados; si se concluye que "A" es mejor que "B", será porque hay un mayor porcentaje de pacientes recuperados que tomaron el medicamento "A". Si se aplica una prueba inferencial adecuada, por ejemplo, una prueba exacta de Fisher, podríamos obtener un valor de probabilidad menor al nivel de error tipo I permitido.

Este valor, conocido como valor  $p$ , es la probabilidad relativa de equivocarse al rechazar una hipótesis que indica que ambos tratamientos son igual de efectivos, como se vio anteriormente, nuestra hipótesis nula; pues bien, si el valor  $p$ , es pequeño con relación al nivel de error tolerado, la probabilidad de equivocarse al rechazar la hipótesis de no diferencia es pequeña, y por tanto decidimos que hay suficiente evidencia para rechazar esa hipótesis y concluimos que en la población de donde provienen esos pacientes que recibieron los tratamientos en mención, el tratamiento "A", parece ser más efectivo que el tratamiento "B". Sin embargo, no será suficiente reportar un valor de probabilidad significativo para creer que uno de los tratamientos es superior al otro, pues esto resuelve parcialmente el problema planteado: habrá que indicar cuál es la *sensibilidad* con la que se calculó este valor  $p$ , en otras palabras indicar qué tan confiable podría ser este valor de probabilidad reportado.

Una manera de indicar qué tan confiable es ese valor es



## ño del efecto

reportar la potencia (calculada a posteriori o la especificada durante el cálculo del tamaño de muestra). La potencia, puede ser definida como la probabilidad de rechazar una hipótesis de no efecto (de igualdad en los efectos), cuando esta es realmente falsa, es decir la capacidad de las pruebas para detectar diferencias entre grupos, cuando estas realmente existen y no por efecto del azar.

El otro concepto que nos importa es, el “tamaño del efecto”: nuestro primer acercamiento fue cuando aprendimos en un curso básico de estadística a calcular el coeficiente de correlación, y si la correlación era igual a 0.89, decíamos que dos variables estaban relacionadas linealmente en una magnitud del 89 %. Por tanto, definiremos al tamaño del efecto como la magnitud de la diferencia de la variable respuesta entre los grupos que fueron comparados en nuestra investigación. Regresando al ejemplo de los tratamientos, el tamaño del efecto será el grado de generalización que posee esa superioridad de “A” sobre “B” en la población de la que se extrajo la muestra. Otra forma de entender el tamaño del efecto, es pensar en qué tanto el valor de una variable respuesta se puede explicar, predecir o controlar por influencia de la variable independiente. Omitir la estimación del tamaño del efecto podría llevar a tomar decisiones fundadas en el desconocimiento de la magnitud de los efectos encontrados; de igual manera si desconocemos el poder de la prueba, se carecerá del conocimiento de conocer errores tipo beta (aquellos que suceden cuan-

do aceptamos una hipótesis nula falsa), que constituyen la prueba de mayor relevancia acerca de la validez del diseño de un estudio, ya que una diferencia significativa no es necesariamente una diferencia grande o importante, para ello habrá que cuantificar la magnitud de dicha diferencia significativa.

No obstante, el reporte de la potencia y el tamaño del efecto en los estudios de investigación no es una práctica habitual entre investigadores, a pesar que desde hace ya más de diez años, algunas revistas y asociaciones científica han insistido en incluir estos estadísticos en la presentación de resultados, por ejemplo la American Psychological Association (APA) ha recomendado incluir el tamaño del efecto y la potencia junto a los valores  $p$ , para la apreciación de la magnitud o importancia de los hallazgos del estudio. De hecho, APA recomienda el uso de intervalos de confianza, exposición de los valores de las medias y desviaciones estándar de cada grupo, el reporte de los valores exactos de probabilidad, la potencia y el tamaño del efecto. Varios investigadores como Ledesma y colaboradores (2008) atribuyen el desuso de la potencia y el tamaño del efecto a que los programas estadísticos más populares no siempre incluyen la estimación del tamaño del efecto entre sus opciones de análisis. Por otro lado, Prajapati y colaboradores (2010), subrayan que la potencia y la estimación del tamaño de muestra (hasta ahora mencionado), es importante porque no resultaría éticamente aceptable conducir un estudio que no tuviese la suficiente capacidad para detectar un

efecto real debido a la pérdida de poder estadístico, ni tampoco resultaría éticamente aceptable conducir un estudio reclutando más participantes de los que fuesen necesarios para detectar una diferencia real.

La potencia estadística se calcula sobre el conocimiento del tamaño de muestra, el nivel de error alfa y el tamaño del efecto. La potencia aumenta si el tamaño de muestra y el tamaño del efecto son mayores, y disminuye si el error alfa disminuye; es por ello que en un estudio analítico no siempre resulta conveniente tomar más muestras de las que se calcularon, pues se corre el riesgo de aumentar la probabilidad de error alfa.

Existen diversos estadísticos del tamaño del efecto como el coeficiente de correlación biserial de Pearson para variables cuantitativas continuas, el coeficiente de contingencia y la  $V$  de Cramer para evaluar la asociación entre dos variables cualitativas, el estadístico  $F$  o el coeficiente  $\eta^2$  cuadrado de un análisis de varianza, el estadístico de Wald de una regresión logística o el estadístico de Wilks de un análisis multivariado de la varianza. También hay otro grupo de estadísticos utilizados en epidemiología como los son los *odds ratios* en estudios de casos y controles, el riesgo relativo de los estudios de cohortes y los *hazard ratios* de los estudios de supervivencia. El saber interpretarlos y requerirlos nos ayudará a valorar de una forma crítica la literatura científica y a no conformarnos con los tradicionales valores de probabilidad, cuya boga fue los años noventa.