

Prueba de hipótesis frente a intervalos de confianza

Carlos Escalante Angulo*

RESUMEN

Los valores P , generalmente, 0,05 ó 0,01, que en la prueba de hipótesis estadística se usan para diferenciar resultados estadísticamente significativos de los no significativos, se considera de poco valor informativo y práctico cuando el investigador biomédico y epidemiólogo están interesados en conocer la magnitud de un resultado de un estudio. Este artículo

muestra la ventaja de los intervalos al comparar la prueba de hipótesis con la estimación de intervalos de confianza para inferir la diferencia entre dos proporciones muestrales.

Palabras clave: valores p , prueba de hipótesis estadística, intervalos de confianza, muestra, población.

* Antropólogo de la Universidad Nacional. Magister en Sociología. Profesor emérito de la Universidad Nacional y la Universidad de La Salle. Correo electrónico: caescalante@unisalle.edu.co.

Fecha de recepción: 10 de enero del 2010
Fecha de aprobación: 8 de abril del 2010

Hypothesis testing versus confidence intervals

ABSTRACT

The p values, generally 0.05 or 0.01, that in the statistical hypothesis test are used to distinguish the significant and non-significant statistical results, it is considered of few information and few practical value, when the biomedical researcher and epidemiologist are interesting in knowing the magnitude of a result of study. This communication shows the

advantage of the intervals to compare the hypothesis test with the interval confidence for inferring the difference between two sample proportions.

Keywords: p -values, statistical hypothesis testing, confidence intervals, sample population.

INTRODUCCIÓN

En décadas pasadas se ha incurrido en conceder excesiva importancia a las pruebas de hipótesis y a los valores p asociados a ellas, para determinar si los resultados de un estudio resultan ser o no ser significativos. Esto ocurre cuando los investigadores están interesados, sobre todo, en la magnitud de las diferencias, en los grupos comparados en los estudios analíticos, sean éstos ensayos clínicos u observacionales. Este excesivo énfasis en el uso de p relega a un segundo plano las pruebas de estimación de intervalos de confianza a las que se les han venido reconociendo mayor valor informativo y utilidad práctica en los estudios clínicos y epidemiológicos.

Es conveniente recordar que el valor p se relaciona con las pruebas de hipótesis y representa la probabilidad de que, en una investigación, un resultado se deba al azar, cuando en realidad no hay relación en la población entre las variables implicadas por ejemplo, una diferencia de proporciones entre dos grupos. En el proceso inferencial se considera que si el valor p es menor que alfa el nivel seleccionado por el investigador, generalmente 0,05 o 0,01 se rechazará la hipótesis nula.

En esta breve nota expondremos críticas autorizadas al uso inapropiado de las pruebas de hipótesis y, a continuación, mostraremos en una tabla resultados de un supuesto estudio y lo analizaremos utilizando sendas pruebas de hipótesis e intervalos de confianza, para mostrar diferencias y virtudes de ambos procedimientos estadísticos. La tabla es la siguiente:

Ejemplo: Resultado: dos muestras independientes

Respuesta	Tratamiento	
	A	B
Con mejoría	80	40
Sin mejoría	20	60
Total	100	100

CRÍTICAS AL USO DE P

Desde hace varias décadas, las críticas autorizadas al uso inapropiado de las pruebas de hipótesis para valorar la importancia de los resultados en investigaciones epidemiológicas y biomédicas han sido recurrentes. Como marco justificativo de este ensayo expositivo, he escogido dos de esas críticas, que expongo a continuación.

Cita 1:

“Adiós, p menor del 0,05”, equivoco y traicionero compañero de viaje. Tus efectos colaterales y toxicidad intracerebral son demasiado grandes para compensar cualquier beneficio que pudieras aportar (Feinstein, 1985).

Cita 2:

Las típicas aseveraciones “ $p < 0,05$, $P > 0,05$, o $P = NS$ ” dan poca información sobre los resultados de un estudio y se basan en el consenso arbitrario de utilizar el nivel de significación estadístico de 5% para definir los posibles resultados: significativo o no significativo. Esto no sirve para nada y, además, favorece la vagancia intelectual. Incluso cuando se indica el valor p en concreto, no se proporciona información alguna sobre las diferencias en los grupos estudiados (Gardner Martin., Altman, 1996)¹.

PRUEBA DE HIPÓTESIS: Z

1. Hipótesis nula (H_0). No hay diferencias en mejoría entre los tratamientos A y B. Hipótesis alterna (H_1): el tratamiento A es más efectivo.
2. Prueba estadística. Prueba Z para la diferencia entre dos proporciones, divida ésta por el error estándar de la diferencia.

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1(1 - P_1)}{n_1} + \frac{P_2(1 - P_2)}{n_2}}}$$

¹ Agradezco al Dr. Juan Manuel Lozano, del Departamento de Epidemiología de la Universidad Javeriana, la gentileza de haberme proporcionado copia de este artículo.

3. Regla de decisión. Sea alfa 0,05, para que el resultado concuerde con 0,95 en que se basa el intervalo de confianza.
4. Valor crítico, 1,96 para la prueba de una cola. Se rechaza H_0 si el valor observado de Z es mayor que 1,96.

Resultado:

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{(0.80)(0.20)}{100} + \frac{(0.40)(0.60)}{100}}} = 6.6$$

Decisión estadística: $6,6 > 1,96$, se rechaza H_0 .

Puesto que en la tabla no existe el valor p más cercano a 6,6, se recomienda reportar $p < 0,001$. Con este nivel de significación se rechaza H_0 y es todo lo que se suele decir: la diferencia entre las proporciones es estadísticamente significativa, cuando el objetivo central de la investigación es demostrar la magnitud de la diferencia en términos de mejoría entre los dos tratamientos. Para observarla hay que recurrir a la estadística descriptiva, $0,80 - 0,40$.

INTERVALO DE CONFIANZA

Como sabemos, la estimación es un procedimiento inferencial por medio del cual se llega a inferir los valores poblacionales por medio de los valores muestrales. La estimación se apoya en la suposición de que el investigador tiene interés en los parámetros de varias poblaciones, tales como medias, proporciones o diferencia entre medias y proporciones. En los estudios analíticos en salud se utiliza la estimación por intervalos, que consiste en dos valores numéricos que definen un intervalo, el cual contiene el parámetro bajo estudio con un nivel de confianza

determinado. Los niveles de confianza (0,90, 0,95, 0,99) están definidos por los correspondientes valores Z (1,64, 1,96, 2,58) que, respectivamente, acompañan el error estándar de la distribución muestral de una estadística. No es ocioso recordar una vez más que el error estándar es una medida de la variabilidad del factor estudiado en una distribución de muestreo, pues se supone de manera implícita que una sola muestra no sirve para determinar con exactitud el valor poblacional. Por lo general, se lo utiliza en el cálculo de los intervalos de confianza.

Ahora calcularemos el intervalo de confianza, con una probabilidad de los 95% para la diferencia de proporciones arriba señaladas, con la siguiente estadística de prueba.

$$IC = (P_1 - P_2) \pm Z(1 - \alpha) \sqrt{\frac{P_1(1 - P_1)}{n_1} + \frac{P_2(1 - P_2)}{n_2}}$$

Expresado en palabras, representa el estimador \pm el nivel de confianza multiplicado por el error estándar de la diferencia:

$$IC = (0.80 - 0.40) \pm Z(1 - \alpha) \sqrt{\frac{(0.80)(0.20)}{100} + \frac{(0.40)(0.60)}{100}} = 0.26 - 0.53$$

El lector ya se habrá dado cuenta de que la amplitud del intervalo y la precisión de la estimación depende del tamaño muestral. Si éste fuera de 200 para cada tratamiento, los límites del intervalo serían 39,25 – 40,75, menor amplitud, mayor precisión.

Una vez más, la estimación del parámetro a partir de los valores muestrales utiliza el error estándar, de modo que en muestras repetidas, el 95% de los intervalos incluiría la diferencia verdadera en la población. En este ejemplo estaría entre 0,26-0,53.

REFERENCIAS

- Anderson, R.P. (1998). *La estadística razonada: reglas y principios* (Cap. 3, 78-79). Barcelona: Paidós.
- Dawson, B. & Trapp, R.O. (2005). *Bioestadística Médica* (Cap. 6). Bogotá: el Manual Moderno.
- Gardner, M.J. & Althman, D.G. (1986). Confidence Intervals Rather than P Values; Estimation Rather than Hypothesis Testing. *British Medical Journal*, 292, 746-750.
- López, L., J.M. & Dennis V., R.J. (2001). Medidas de frecuencia, de asociación y de impacto. En A. Ruiz Morales, C. Gómez Restrepo & D. Londoño Trujillo (Edits. académicos). *Investigación clínica: epidemiología clínica aplicada* (93-109). Bogotá: Centro Editorial Javeriano.
- Prieto, V., L. & Herrán, I. (2005). *Qué significa ¿Estadísticamente significativo?* Madrid: Edición Díaz Santos.