El supuesto de normalidad: ¿mito o realidad?

Myrian Vergara* / Giovany Babativa**

Resumen

En este artículo se presenta una breve reseña del desarrollo de la teoría de pruebas de hipótesis y se aclaran las principales diferencias entre pruebas paramétricas y no paramétricas. Para ilustrar el uso de unas y otras, se realiza un estudio de simulación, en el que se muestran los errores en que incurren los investigadores al tomar decisiones cuando pasan por alto los supuestos exigidos por un conocido método estadístico.

Palabras clave: Estadística paramétrica, estadística no paramétrica, prueba t-student, prueba del signo.

The assumption of normality: myth or reality?

Abstract

In this article presents a brief overview of the development of the theory of hypothesis testing, and clarified the main differences between parametric and non-parametric tests. To illustrate the use of other tests, conducting a study of simulation shown errors incurred by researchers to make decisions when ignored the assumptions required by a well-known statistical method.

Keywords: parametric statistics, Non-parametric statistical, test t-student, test sign.

* Docente del Departamento de Ciencias Básicas de la Universidad de La Salle.
Correo electrónico: mvergara@unisalle.edu.co

** Docente del Departamento de Ciencias Básicas de la Universidad de La Salle.
Correo electrónico: jobabativa@unisalle.edu.co

Fecha de recepción: 23 de marzo de 2010
Fecha de aprobación: 1 de junio de 2010

Introducción

La estadística, como parte fundamental de las ciencias básicas, es una herramienta que le permite al científico describir la naturaleza de sus investigaciones con ayuda de mediciones cualitativas o cuantitativas. Más aún cuando las mediciones se hacen sobre muestras aleatorias de los objetos observados, el científico puede generar modelos, inferencias y predicciones al conjunto de dichos objetos.

Muchos investigadores se han enfrentado con el hecho de seleccionar técnicas de inferencia adecuadas que les permitan interpretar los datos como evidencia a favor o en contra de hipótesis formuladas previamente. Estos métodos de contraste requieren ciertos supuestos teóricos, que muchas veces no se satisfacen o no son fácilmente verificables. Por tal motivo, nuestro objetivo es mostrarle al lector la importancia de los supuestos para la toma de decisiones y sugerir métodos estadísticos alternativos cuando éstos no se cumplen.

Breve referencia histórica

Para hacer una breve reseña histórica sobre las pruebas de hipótesis, nos remitimos al año 1735, en el que Daniel Bernoulli (1700-1782) propuso probar la hipótesis de aleatoriedad de las inclinaciones de las órbitas planetarias. Posteriormente, en 1900, Karl Pearson (1857-1936), considerado el fundador de la estadística como ciencia, desarrolló el coeficiente de correlación y la prueba de bondad de ajuste chi-cuadrado. Sin embargo, la formulación de pruebas de hipótesis como se enseñan y se emplean hoy en día fue desarrollada entre 1915 y 1933 por los ingleses Ronald Fisher (1890-1962), Egon Pearson (1985-1980) y el polaco Jerzy Neyman (1894-1981).

En diferentes textos y artículos de la historia de la estadística se mencionan las diferencias personales que existieron entre Fisher y Pearson. A este último, fundador y director de la revista *Biometrika* (revista que en la actualidad cuenta con gran prestigio internacional), no le agradaban las demostraciones de Fisher y se dedicó a ignorar su trabajo, llegando inclusive a rechazarle valiosas publicaciones. Esta situación creó una enemistad que llegó a extenderse a E. Pearson hijo, K. Pearson y Neyman, su compañero de fórmula.

Si bien actualmente se enseña la aplicación de pruebas de hipótesis como una metodología unificada, es interesante saber la historia oculta que existe detrás de esta serie de pasos, que no son más que un híbrido resultante entre dos enfoques diferentes de la inferencia estadística, realizados separadamente por Fisher, por un lado y, Neyman y E. Pearson por el otro.

En el enfoque de Fisher, la hipótesis nula H_0 indica que la muestra proviene de una población infinita, con distribución conocida que depende de algún parámetro desconocido θ. Dicha hipótesis se rechaza si la estimación muestral del parámetro desconocido es muy diferente de este; el rechazo de H_0 se hace con una probabilidad pequeña de error llamada valor p que es la mínima probabilidad con que la hipótesis nula se rechaza con la evidencia que da la información de la muestra. Por depender de la muestra, p es una estadística con distribución uniforme en el intervalo (0,1). Esto es, probar por ejemplo que la hipótesis nula $H_0 \colon \theta = \theta_0$ con θ_0 conocido implica usar la información contenida en la muestra como evidencia inductiva en contra de la hipótesis nula.

Sin embargo, Fisher hace notar que al establecer algún criterio para decidir entre rechazar o no la hipótesis nula, el investigador podría cometer un error en su decisión, llegar a rechazar la hipótesis nula cuando ella es verdadera, lo cual se conoce como nivel de significación estadística.

A diferencia de Fisher, E. Pearson y Neyman se dedicaron a elaborar una teoría con el objetivo de obtener pruebas óptimas de significación. Definen dos hipótesis complementarias: la hipótesis nula H_0, que abarca una parte de los valores de θ, la cual se compara con una hipótesis alternativa H_1, que cubre los restantes valores posibles de θ. Esta formulación permite caer en un segundo tipo de error, que consiste en no rechazar una hipótesis nula cuando es falsa. Fueron Pearson y Neyman quienes, al introducir un segundo tipo de error (denotado por β), denominaron error tipo uno (denotado por α) al error introducido por Fisher. A partir del error β, introdujeron el concepto de "potencia de una prueba", definida como 1-β, la cual se refiere a la capacidad que tiene la prueba de detectar cuándo la hipótesis nula es falsa. Esta teoría sustituye el valor p, por una regla de decisión basada en el nivel de significación de la prueba.

La teoría de Pearson y Neyman establece unas reglas de comportamiento inductivo, en contraposición a la inferencia inductiva de Fisher [3]. De manera que se decide en favor de una u otra hipótesis, sin considerar la intuición personal que el investigador tenga sobre una de ellas.

Urbisaia y Brufman (2005) describen que entre 1954 y 1960, el problema fue resuelto silenciosamente en los libros de texto escritos por no estadísticos en su afán de enseñar a sus estudiantes "reglas de la estadística".

Estudio de simulación

Los métodos de probar hipótesis suelen clasificarse de acuerdo con el tipo de supuestos que se hacen acerca de la distribución de la población de la que proviene la muestra (distribución muestreada). Distinguiremos aquí dos de estos métodos, conocidos como pruebas paramétricas y pruebas no paramétricas. Se llaman paramétricas aquellas pruebas para cuya aplicación se requiere el conocimiento de la forma funcional de la distribución muestreada y se llaman pruebas no paramétricas aquellas para las que sólo exigen supuestos como la continuidad o la simetría de la distribución muestreada. En estas condiciones, las pruebas más adecuadas para contrastar las hipótesis de interés fueron desarrolladas por Pearson y Neyman (Lema de Neyman-Pearson).

Entre las pruebas paramétricas más conocidas y aplicadas tenemos la prueba *t*-student utilizada para contrastar hipótesis acerca de la media y la prueba F utilizada para contrastar hipótesis acerca de la varianza, las cuales exigen que la distribución muestreada sea la normal. Entre las pruebas no paramétricas más utilizadas se encuentran la prueba del signo, la del rango signado de Wilcoxon, la de Mann-Whitney-Wilcoxon y la de Kruskal-Wallis, utilizadas para contrastar hipótesis acerca de la mediana en una, dos y k muestras, respectivamente.

A continuación se presenta un caso en el que los datos no siguen una distribución normal y se muestra el impacto en las decisiones cuando se aplican métodos paramétricos.

Es común encontrar artículos, tesis o trabajos en los que se aplique la prueba *t*-student para contrastar hipótesis de medias sin que se comprueben los supuestos requeridos. Al parecer muchos investigadores consideran que el supuesto según el cual los datos provienen de una distribución normal no es más que un mito. Sin embargo, la construcción de la estadística de prueba está basada en que si $X_1, \ldots, X_n$ son variables aleatorias con distribución normal con media μ y varianza σ^2 entonces:

$$Z_n = \frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Y:

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

Luego,

$$T = \frac{\overline{X_n} - \mu}{S/\sqrt{n}} = \frac{\frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{S^2}{\sigma^2}}} = \frac{\frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}\frac{1}{n-1}}} = \frac{Z_n}{\sqrt{\frac{X^2}{n-1}}} \sim t_{(n-1)}$$

Fisher demostró que el cociente entre una variable normal estándar Z_n y la raíz cuadrada de una variable chi-cuadrado X^2, sobre sus grados de libertad (n-1), sigue una distribución *t*-student con los grados de libertad de la chi-cuadrado.

Nótese que todo parte del supuesto según el cual los datos originales provienen de una distribución normal, de modo que es apenas natural realizar la pregunta ¿qué sucede si aplico la prueba *t*-student a datos que no tienen una distribución normal? Para resolver esta inquietud, a continuación se muestran los resultados de una simulación.

Suponga que desea contrastar el siguiente sistema de hipótesis:

$$H_0:\ \mu = 0\ vs.\ H_1:\ \mu \neq 0$$

Si pensamos en una estadística paramétrica, es natural que recurramos a la estadística *t*-student, definida por:

$$T = \frac{\overline{X_n} - \mu}{S/\sqrt{n}}$$

Ahora, existen otras estadísticas basadas en métodos no paramétricos que podrían ser utilizadas, una de ellas es la estadística del signo, definida por:

$$S = \sum_{i=1}^{n} I(X_i > 0)$$

Donde $I(X_i > 0) = 1\ \ si\ X_i > 0\ o\ 0\ si\ X_i \leq 0$ es el número de observaciones positivas.

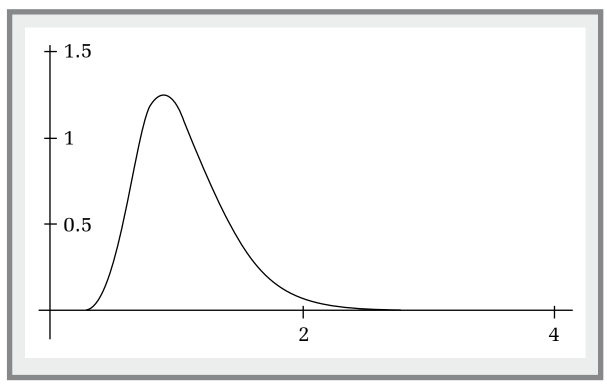
La simulación consiste en:

- Generar 1.000 muestras independientes de tamaño n de una distribución Lognormal con media igual a cero (figura 1), es decir, que la hipótesis nula es verdadera.

- Calcular el valor de la estadística *t*-student y de la estadística del signo.

- Determinar en cada caso si la hipótesis nula es rechazada.

Figura 1. Función de densidad de una lognormal con $\mu = 0$ y $\sigma^2 = 1/3$.



Utilizando un nivel de significación del 5%, y de acuerdo con la teoría estadística, el investigador esperaría que el porcentaje de veces que la prueba se equivoque sea muy cercano al nivel nominal (5%).

Se probaron diferentes tamaños de muestra $n = 50;$ 100; 200; 300 y 500, y se determinó el error de tipo I (que la prueba rechace la hipótesis nula cuando no debería hacerlo), los resultados se muestran en la tabla 1.

Tabla 1. Estimación del error de tipo I.

Tamaño de muestra	Porcentaje de rechazos	
	Prueba T	Prueba S
50	19,1%	5,3%
100	37,4%	5,2%
200	66,8%	5,4%
300	84,0%	5,0%
500	96,2%	4,9%

Se puede observar que la prueba *t*-student se equivoca un porcentaje de veces superior al esperado 5%, esto es consecuencia de que los datos originales no provienen de una distribución normal. Además, el porcentaje de veces que se equivoca incrementa en la medida que el tamaño de muestra es más grande, pues con un tamaño de muestra de 50 se equivoca en un 19,1% de la veces, mientras que con un tamaño de muestra de 500 se equivoca en el 96,2% de las veces. Por otro lado, obsérvese que la prueba del signo se mantuvo alrededor del nivel nominal del 5%, convirtiéndose en la prueba más adecuada cuando los datos no siguen una distribución normal.

Conclusiones

Es importante que los investigadores, cuando tengan este tipo de hipótesis, y en general cualquier procedimiento estadístico, comprueben los supuestos necesarios, pues como se observó en la tabla 1 el investigador puede caer en falsas conclusiones cuando no se tienen en cuenta los requerimientos mínimos exigidos por un método estadístico.

Los interesados en el programa de simulación pueden solicitarlo vía correo electrónico a los autores.

Referencias bibliográficas

Gibbons, J. y Chakraborti, S. (2006). *Nonparametric statistic inference*. Cuarta edición. Nueva York: Marcel Dekker.

Figueroa, G. (2002). Ronald Aylmer Fisher. *Apuntes de la Historia de las Matemáticas*. Vol. 3. N.° 1.

Rodríguez, E. (2005). Estadística y psicología: análisis histórico de la inferencia estadística. *Revista Electrónica PsicologiaCientifica.com*.

Urbisaia, H. y Brufman, J. (2008). La controversia entre Fisher y Neyman-Pearson sus implicaciones en la investigación econométrica. XIV Jornadas de Epistemología de las Ciencias Económicas. Universidad de Buenos Aires, Facultad de Ciencias Económicas, 2 y 3 de octubre de 2008. Extraído desde http://www.econ.uba.ar/www/institutos/epistemologia/marco_archivos/XIV%20Jornadas%20de%20Epistemologia/Jornadas/ponencias/Actas%20XIV/Trabajos%20Episte/UrbisaiaBrufman.pdf

Zaven A. y Dudewicz, E. (2000). *Fitting statistical distributions. The generalized lambda distribution and generalized bootstrap methods.* Ohio: Chapman & Hall.