Revista Mexicana de Cardiología

Comunicado breve

Volumen 26, Número 1
Enero - Marzo 2015
pp 59 - 61

t-Student. Usos y abusos

Reinaldo Alberto Sánchez Turcios*

RESUMEN

La prueba t-Student se fundamenta en dos premisas; la primera: en la distribución de normalidad, y la segunda: en que las muestras sean independientes. Permite comparar muestras, N ≤ 30 y/o establece la diferencia entre las medias de las muestras. El análisis matemático y estadístico de la prueba con frecuencia se minimiza para N > 30, utilizando pruebas no paramétricas, cuando la prueba tiene suficiente poder estadístico.

Palabras clave: t-Student, distribución de normalidad, estadística.

ABSTRACT

Student's t test is based on two premises; first: normality of distribution and second: the independence of the samples. This allows comparing samples N ≤ 30 and/or establishes the differences between the means of the two samples. The mathematical and statistical analysis of the test is frequently minimalized N > 30, using non parametric tests, when the test has enough statistical power.

Key words: *Student's t, normal distribution, statistic.*

INTRODUCCIÓN

Con el seudónimo de estudiante (Student), William Sealy Gosset desarrolló la prueba t y la distribución t.[1] Esta prueba se usa con frecuencia en las publicaciones médicas indexadas nacionales e internacionales y se han observado errores consistentes (*The New England Journal of Medicine*, *Lancet* y *British Medical Journal)*.[2]

El objetivo de esta comunicación es plantear correctamente la prueba y distribución t. La distribución t es un conjunto de curvas estructurada por un grupo de datos de unas muestras en particular. La contribución de esta prueba, específicamente, es para comparar dos muestras de tamaño ≤ 30. La primera presunción es formular la hipótesis nula y la hipótesis alterna, que establece que no hay diferencias en la media de las dos muestras independientes y que de existir esta diferencia, sólo se debe al azar.[3] Si la t calculada que se origina de las dos muestras es desmesurada (valor de p que se encuentra en las tablas respectivas), entonces se rechazaría la hipótesis nula (error tipo I). Es importante mencionar que este valor depende del valor de significancia establecido con anterioridad de lo que se quiere probar,[4] para la diferencia entre las medias de las dos muestras. Este valor de significancia es la probabilidad de rechazar erróneamente la hipótesis nula.

NATURALEZA DE LA t-STUDENT

La t de Student, inicialmente se diseñó para examinar las diferencias entre dos muestras independientes y pequeñas que tengan distribución normal y homogeneidad en sus varianzas (en el artículo original, el autor no define qué es una muestra grande y/o pequeña). Gosset hace hincapié en la normalidad de las dos muestras como crucial en el desarrollo de la prueba.

METODOLOGÍA DE LA t-STUDENT

1. Probar que cada una de las muestras tiene una distribución normal;[‡] 2. Obtener para cada una de las muestras: a) el tamaño de las muestras (n_1 y n_2), b) sus respectivas medias (m_1 y m_2), c) sus varianzas (v_1 y v_2); 3. Probar que las varianzas sean homogéneas;

* UMAE Hospital de Cardiología, Centro Médico Nacional Siglo XXI, IMSS.

Este artículo puede ser consultado en versión completa en http://**www.medigraphic.com/revmexcardiol**

‡ La t-Student es una prueba poderosa, en la que aunque una de las muestras no tenga distribución normal pero la otra sí y la razón de la varianza más grande a la más pequeña sea < 2, esta prueba resulta adecuada al comparar dos medias.

‡‡4. En caso de homogeneidad en esas varianzas: a) establecer la diferencia entre las medias: m_1-m_2, b) calcular la varianza común de las dos muestras.

$$vc = ((n_1-1)v_1 + (n_2-1)v_2)/(n_1 + n_2 - 2)$$

Es decir, la varianza común (vc) es igual a un promedio pesado de las varianzas de las dos muestras en donde los pesos para ese promedio son iguales al tamaño, menos uno (n-1) para cada una de las muestras, c) con esa varianza común, se calcula el error estándar de la diferencia de las medias ESM= $\sqrt{((vc)\,(n_1 + n_2)/(n_1 n_2))}$; 5. Finalmente, la t-Student es igual al cociente de la diferencia de medias entre el ESM anterior; 6. De acuerdo con nuestra hipótesis nula y alterna se debe demostrar que existe diferencia entre las medias de las muestras, se consulta una tabla de t-Student con grado de libertad igual a n1 + n2-2 y se calcula el valor de P.[5]

Ejemplo

De un universo de 44,000 niños, a los que se les registró el peso, talla e índice de masa corporal, se tomó una muestra de 56 adolescentes (21 niñas y 35 niños), del subgrupo de niñas y niños de 14 años de edad, para comparar las medias tomando exclusivamente el índice de masa corporal (IMC).

IMC en niñas y niños de 14 años de edad

Paso 1: prueba de normalidad de cada una de las muestras.

N	$\bar{X}$	S	V
Niñas 21	21.775	4.225	17.852
Niños 35	20.850	3.798	14.428

Niñas: p = 0.071, hay normalidad.
Niños: p = 0.0008 no hay normalidad.

Paso 2: en este caso se hace la prueba t-test aun sabiendo que una de las muestras (los niños) no tiene normalidad.

‡‡ En este caso se debe utilizar una modificación a la t-Student dada por Satterthwaite y Welch.[6,7]

Paso 3: prueba para la homogeneidad de varianzas; se pueden considerar que son homogéneas debido a que la p = 0.570.

Paso 4: (i) diferencia de medias = 0.025, (ii) vc a las muestras.

$$vc = ((n_1-1)v_1 + (n_2-1)v_2)/(n_1 + n_2 - 2)$$

$$= \frac{20 \times 17.852 + 34 \times 14.428}{21 + 35 - 2}$$

$$= \frac{357.040 + 490.552}{54}$$

$$= \frac{847.592}{54}$$

$$=15.696$$

(iii) Error estándar de las diferencias de las medias

$$ESM = \sqrt{(vc)\,\frac{n_1 + n_2}{(n_1)\,(n_2)}}$$

$$=\sqrt{(15.696)\,((56)/(735))}$$
$$=\sqrt{(15.696)\,(0.076)}$$
$$=\sqrt{1.196}$$
$$= 1.094$$

Paso 5: el valor de la t-test será:

t = (diferencia de las medias)/(ESM)

$$t = \frac{0.925}{1.094}$$

$$t = 0.846$$

Paso 6: hipótesis:

Ho: el IMC es igual en niños y niñas.
H1: El IMC es diferente entre los niños y las niñas.

Los grados de libertad, para consultar la tabla de t-Student son 21 + 35-2 = 54, consultando el valor de p es 0.401.

Por lo tanto, no existe diferencia entre el IMC entre los niños y niñas de 14 años.

COMENTARIO

En la masa crítica de información científica, principalmente en el área médica, cuando no hay normalidad en ambas muestras se pretende sustituir a esta prueba por la U de Mann-Whitney, pero puede ser cierta esta sustitución cuando las muestras tienen distribución similar y cuando una curva está desplazada con respecto a la otra y sólo en estos casos darían resultados verosímiles a la diferencia entre las dos muestras. Desde otra óptica se ha usado estadística no paramétrica en muestras grandes cuando la t-Student (si se cumplen las condiciones) daría mejor resultado que el uso de las pruebas anteriores.

CONCLUSIONES

1. Es necesario mencionar que la distribución t-test es similar a la distribución de Gauss cuando las muestras > 30.
2. El poder estadístico tiene mayor magnitud cuando las condiciones que se necesitan lo cumplen ambas muestras, independientemente del tamaño.
3. La prueba original demostró que existe una curva que describe el comportamiento de la diferencia de medias y permite calcular el área bajo la curva que representa la probabilidad de la diferencia entre ellas.

BIBLIOGRAFÍA

1. By Student. The probable error of a mean. Biometrika. 1908; 6: 1-25.
2. Fagerland MW. t-tests, non-parametric tests, and large studies-a paradox of statistical practice? BMC Med Res Methodol. 2012; 12: 78-85.
3. Dawson-Saunders B, Trapp Robert G. Bioestadística Médica. México, Editorial Manual Moderno, 1993.
4. Wayne W. Daniel. Bioestadística base para el análisis de las ciencias de la salud. 4ª ed. México, Limusa Wiley. 2002.
5. Zar Jerrold H. Biostatistical analysis. Prentice Hall. Inc. Fifth edition. New York, USA, Prentice Hall, 2010.
6. Satterthwaite FE. An approximate distribution of estimates of variance components. Biometrics Bul. 1946; 2: 110-114.
7. Welch BL. Generalization of "student's" problem when several different population variances are involved. Biometrika. 1947; 34: 28-35.

Dirección para correspondencia

Dr. Reinaldo Alberto Sánchez Turcios
Tepic Núm. 113-610,
Col. Roma Sur,
Del. Cuauhtémoc, México, D.F. 06760
E-mail: rturcios@live.com.mx