

PAPEL DE DISTRIBUCIÓN «T» DE STUDENT EN LOS CONTRASTES DE HIPÓTESIS

González Muñoz, P.
Tardáguila García, P.
Díaz Leno, M.S.
Martín Rodríguez, J.

INTRODUCCIÓN

Cuando es conocida la varianza poblacional, la distribución de la media muestral es $N(m, \frac{s}{\sqrt{n}})$, pero, sin embargo, cuando la varianza poblacional es desconocida, el conocimiento anterior no sirve de nada práctico, puesto que con él no se pueden calcular probabilidades de sucesos correspondientes a la media muestral, y es preciso entonces recurrir a una distribución independiente de la varianza poblacional; ésta distribución es la distribución t de Student.

Es conocido que si tenemos $(n+1)$ variables aleatorias $N(0, s)$ e independientes $(h, h_1, h_2, \dots, h_n)$, la variable t de Student con n grados de libertad se define como

$$t_n = \frac{h}{\sqrt{\frac{h_1^2 + \dots + h_n^2}{n}}}$$

y que no depende de la varianza poblacional s^2 pues

$$\sqrt{\frac{h_1^2 + \dots + h_n^2}{n}} = \sqrt{\frac{\sigma^2}{n} \chi_n^2}$$

y entonces como h es $N(0, s)$, se verifica que $h = s \cdot \chi$, con χ variable $N(0,1)$; así

$$t_n = \frac{h}{\sqrt{\frac{h_1^2 + \dots + h_n^2}{n}}} = \frac{s \chi}{\sqrt{\frac{\sigma^2}{n} \chi_n^2}} = \frac{\chi}{\sqrt{\frac{\chi_n^2}{n}}}$$

lo que nos indica que la distribución t de Student con n grados de libertad se puede definir como el cociente entre una variable $N(0,1)$, y la raíz cuadrada de una variable chi-cuadrado con n grados de libertad, dividida por sus grados de libertad.

DISTRIBUCIÓN DE LA MEDIA MUESTRAL

Con todo esto, si la media muestral, \bar{x} , se distribuye como una variable $N(m, \frac{s}{\sqrt{n}})$ entonces $\bar{x} = m + \frac{s}{\sqrt{n}} \cdot x$ siendo x una variable $N(0,1)$, y así, x sigue una distribución $N(0, s)$.

Como además la variable aleatoria nS^2 es igual a la suma de cuadrados de $(n-1)$ variables $N(0,s)$ e independientes, se verifica entonces que la variable

$$\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{nS^2}{n-1}} = \sqrt{\frac{\chi_{n-1}^2}{n-1}}$$

que es independiente de la media muestral, y es el denominador de la distribución t con $(n-1)$ grados de libertad; entonces la variable

$$\frac{\sqrt{n}(\bar{x} - \mu)}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}}$$

se distribuye como una variable t de Student con $(n-1)$ grados de libertad, es decir,

$$\frac{(\bar{x} - \mu)}{s / \sqrt{n}} \approx t_{n-1}$$

DISTRIBUCIÓN DE LA DIFERENCIA DE MEDIAS MUESTRALES

Si tenemos dos muestras provenientes de una misma población $N(\mu, s)$. La primera muestra (x_1, \dots, x_n) de tamaño n, y la segunda (y_1, \dots, y_m) de tamaño m. Si \bar{x} , e \bar{y} , son las medias muestrales, y las varianzas muestrales respectivamente entonces la variable nS_x^2, mS_y^2 es la suma de cuadrados de $n+m-2$ variables aleatorias $N(0, s)$ e independientes, con lo cual la variable

$$\frac{nS_x^2 + mS_y^2}{\sigma^2}$$

sigue una distribución chi-cuadrado con $n+m-2$ grados de libertad. Además como la variable $\bar{x} - \bar{y}$, sigue una distribución $N(0, \frac{\sigma^2}{nm})$, entonces

$\bar{x} - \bar{y} = \frac{\sigma}{\sqrt{nm}}x$, con x variable $N(0,1)$; y así, la variable $\frac{\sqrt{nm}(\bar{x} - \bar{y})}{\sigma}$ sigue una distribución $N(0, 1)$, y como la variable

$$\sqrt{\frac{nS_x^2 + mS_y^2}{n+m-2}}$$

es independiente de $\bar{x} - \bar{y}$, así la variable

$$\frac{\sqrt{\frac{nm}{n+m}}(\bar{x} - \bar{y})}{\sqrt{\frac{nS_x^2 + mS_y^2}{n+m-2}}}$$

sigue una distribución t de Student con $n+m-2$ grados de libertad.

Si en lugar de tener una única población, se parte de dos muestras provenientes de dos poblaciones con distintas medias, distribuidas $N(\mu_1, s)$ y $N(\mu_2, s)$ respectivamente, entonces la distribución de la diferencia de medias muestrales

$$\frac{\sqrt{\frac{nm}{n+m}}[(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)]}{\sqrt{\frac{nS_x^2 + mS_y^2}{n+m-2}}}$$

sigue una distribución t de Student con $n+m-2$ grados de libertad.

Si las varianzas de las dos poblaciones son diferentes, el resultado anterior no es válido, pero Welch sugirió en 1.937 que el estadístico

$$\frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$$

se distribuye, aproximadamente, como una distribución t de Student con f grados de libertad, siendo

$$f = \frac{\left(\frac{s_x^2}{n} + \frac{s_y^2}{m}\right)^2}{\frac{\left(\frac{s_x^2}{n}\right)^2}{n+1} + \frac{\left(\frac{s_y^2}{m}\right)^2}{m+1}} - 2$$

aunque existen otras aproximaciones, como puede verse en DUDEWICZ (1988).

Todo esto se puede aprovechar en los intervalos de confianza, y en los contrastes de hipótesis, como puede observarse en los dos ejemplos expuestos.

Contraste de hipótesis para la media de una población normal con varianza poblacional desconocida.

Si el tamaño de la muestra es pequeño ($n < 30$), sabemos de párrafos anteriores que la variable

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n-1}}}$$

se aproxima por una distribución t de Student, con $n-1$ grados de libertad, por lo que ésta variable t puede utilizarse con un estadístico de contraste.

La región crítica, en éste caso, estará determinada por los valores de la variable t que sean excesivamente grandes en valor absoluto, si el contraste es bilateral, y para especificar el nivel de cuando pueden considerarse grandes, teniendo en cuenta la distribución de la variable t , que sigue una distribución t con $n-1$ grados de libertad, serán aquellos valores mayores que $t_{\alpha/2, n-1}$ en el contraste bilateral (ver figura 1), o que $t_{\alpha, n-1}$ en el contraste unilateral (ver figura 2).

Las regiones críticas serán pues:

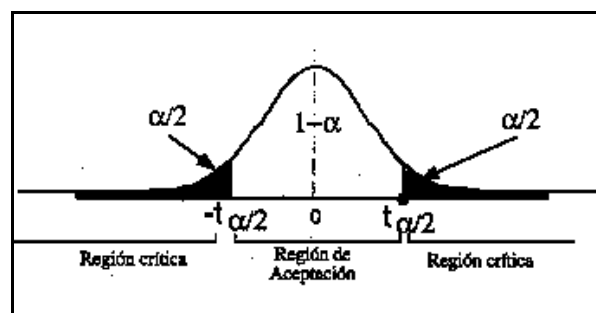


Figura 1. Regiones crítica y de aceptación en el contraste bilateral cuando la variable se distribuye según una t de Student.

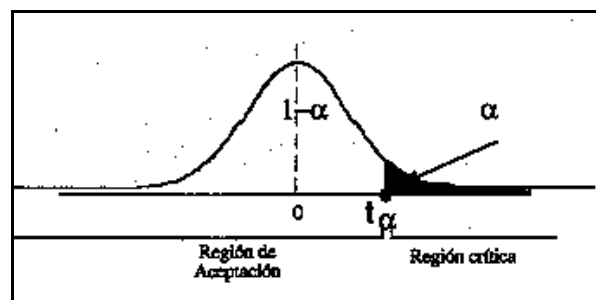


Figura 2. Regiones crítica y de aceptación en el contraste unilateral cuando la variable se distribuye según una t de Student.

CONTRASTE DE HIPÓTESIS PARA LA IGUALDAD DE MEDIAS DE DOS POBLACIONES NORMALES CON VARIANZAS POBLACIONALES DESCONOCIDAS, PERO IGUALES.

Supongamos ahora que las varianzas son desconocidas pero iguales ($s_1 = s_2 = s$). Si las muestras tienen tamaño pequeños, aunque no se conozca la varianza poblacional, se trabaja como si se conociese utilizando en lugar de la varianza poblacional, su estimador la cuasivarianza muestral, por lo que la distribución de la diferencia de medias muestrales, si las muestras son pequeñas ($n+m < 30$), entonces la variable siguiente

$$t = \frac{[(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)]}{\hat{S} \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

$$\hat{S} = \sqrt{\frac{(n-1)\hat{S}_x^2 + (m-1)\hat{S}_y^2}{n+m-2}}$$

sigue una distribución t de Student con $n+m-2$ grados de libertad.

Si la hipótesis nula es cierta, el estadígrafo de contraste que utilizaremos es

$$t = \frac{(\bar{x} - \bar{y})}{\hat{S} \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

porque se distribuye como una t de Student con $n+m-2$ grados de libertad, siendo la región crítica determinada por los valores de ésta variable t, que son mayores en valor absoluto que $t_{\alpha/2}$ en el contraste bilateral, o bien los valores de t, que son mayores en valor absoluto que t_{α} en el contraste unilateral.

Textos básicos en los cuales nuestros alumnos pueden encontrar éste tema detallado y con ejercicios son:

BIBLIOGRAFÍA

- GALINDO, M. P.(1984): *Exposición intuitiva de métodos estadísticos: Fundamentos y aplicaciones a Biología, Medicina y otras ciencias..* Universidad de Salamanca.
- LLOPIS, J. (1996): *La Estadística: una orquesta hecha instrumento.* Barcelona, Ariel.
- MARTIN, A Y LUNA, J. D. (1990): *Bioestadística para las ciencias de la salud.* Madrid, Norma.
- MARTIN , F. J. Y RUIZ-MAYA, L (1995): *Estadística I: Probabilidad.* Madrid, AC.
- ROHATGI, V.K.(1976): *An Introduction to Probability theory and Mathematical Statistics.* New York, Wiley.
- RUIZ-MAYA, L Y MARTIN , F. J. (1995): *Estadística II Inferencia.* Madrid, AC.