Distribución de t Student

Teoría de pequeñas muestras

En probabilidad y estadística, la **distribución-*t*** o **distribución t de Student** es una distribución de probabilidad que surge del problema de estimar la media de una población normalmente distribuida cuando el tamaño de la muestra es pequeño.

A la teoría de pequeñas muestras también se le llama teoría exacta del muestreo, ya que también la podemos utilizar con muestras aleatorias de tamaño grande.

Veremos un nuevo concepto necesario para poder entender la distribución t Student. Este concepto es "grados de libertad".

Para definir grados de libertad se hará referencia a la varianza maestral:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}$$

Esta fórmula está basada en *n-1* ***grados de libertad***. Esta terminología resulta del hecho de que si bien s^2 está basada en *n* cantidades $x_1 - \overline{x}, x_2 - \overline{x}, \ldots x_n - \overline{x}$, éstas **suman cero**, así que especificar los valores de cualquier *n-1* de las cantidades determina el valor restante.

Por ejemplo, si *n=4* *y* $x_1 - \overline{x} = 8$; $x_2 - \overline{x} = -6$ y $x_4 - \overline{x} = -4$, entonces automáticamente tenemos $x_3 - \overline{x} = 2$, así que sólo tres de las cuatro medidas de $x_i - \overline{x}$ están libremente determinadas, la otra debe tomar el valor que haga esta suma cero; es por esto que solo tenemos 3 grados de libertad.

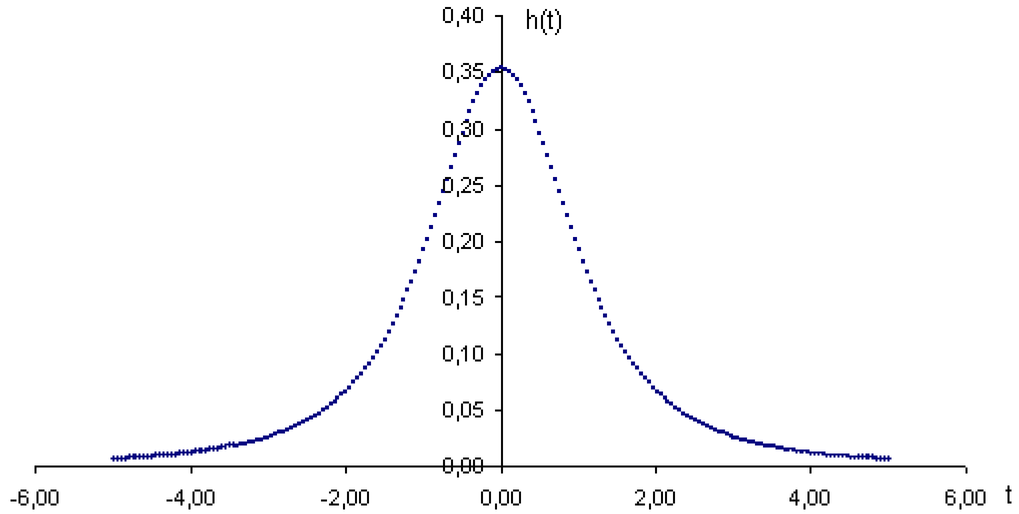
grados de libertad=número de mediciones-1

Distribución de probabilidad t-Student

Una variable aleatoria se distribuye según el modelo de probabilidad **t o T de Student con k grados de libertad**, donde **k** es un entero positivo, si su función de densidad es la siguiente:

$$h_k(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}}\left(1+\frac{t^2}{k}\right)^{-\frac{(k+1)}{2}} \qquad -\infty < t < \infty, \qquad donde \qquad \Gamma(p) = \int_0^{\infty} e^{-x}x^{p-1}dx$$

La gráfica de esta función de densidad es simétrica, respecto del eje de ordenadas, con independencia del valor de **k**, y de forma algo semejante a la de una distribución normal:



Distribución t de Student con 10 grados de liberta

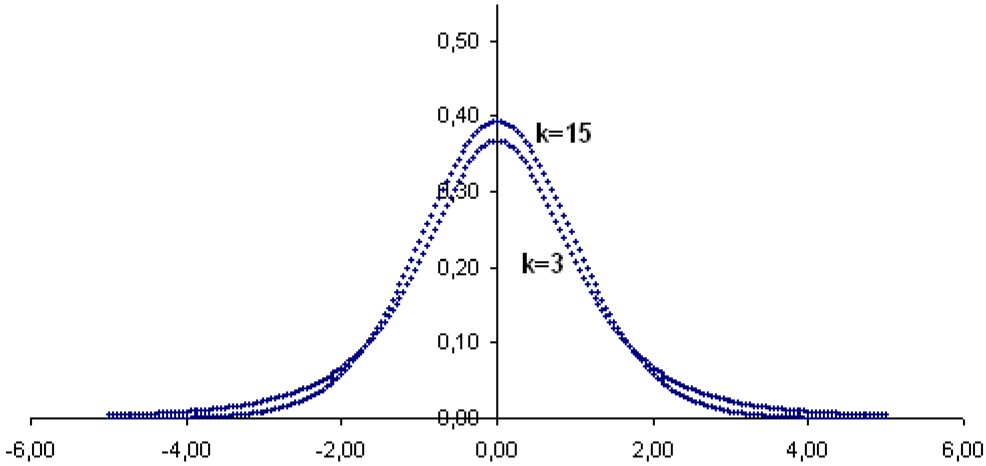
Su valor medio y varianza son

$$E(T) = \mu = \int_{-\infty}^{\infty} t..h_k(t).dt = \int_{-\infty}^{\infty} t.\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}}\left(1+\frac{t^2}{k}\right)^{-\frac{(k+1)}{2}} dt = ... = 0$$

Si k>3

$$Var(T) = \sigma^2 = E((T-\mu)^2) = \int_{-\infty}^{\infty} (t-\mu)^2.h_k(t).dt = \int_{-\infty}^{\infty} t.\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}}\left(1+\frac{t^2}{k}\right)^{-\frac{(k+1)}{2}} dt = ... = \frac{k}{k-2}$$

La siguiente figura presenta la gráfica de varias distribuciones *t*. La apariencia general de la distribución *t* es similar a la de la distribución normal estándar: ambas son simétricas y unimodales, y el valor máximo de la ordenada se alcanza en la media μ = 0. Sin embargo, la distribución *t* tiene colas más amplias que la normal; esto es, la probabilidad de las colas es mayor que en la distribución normal. A medida que el número de grados de libertad tiende a infinito, la forma límite de la distribución *t* es la distribución normal estándar.



Propiedades de las distribuciones t

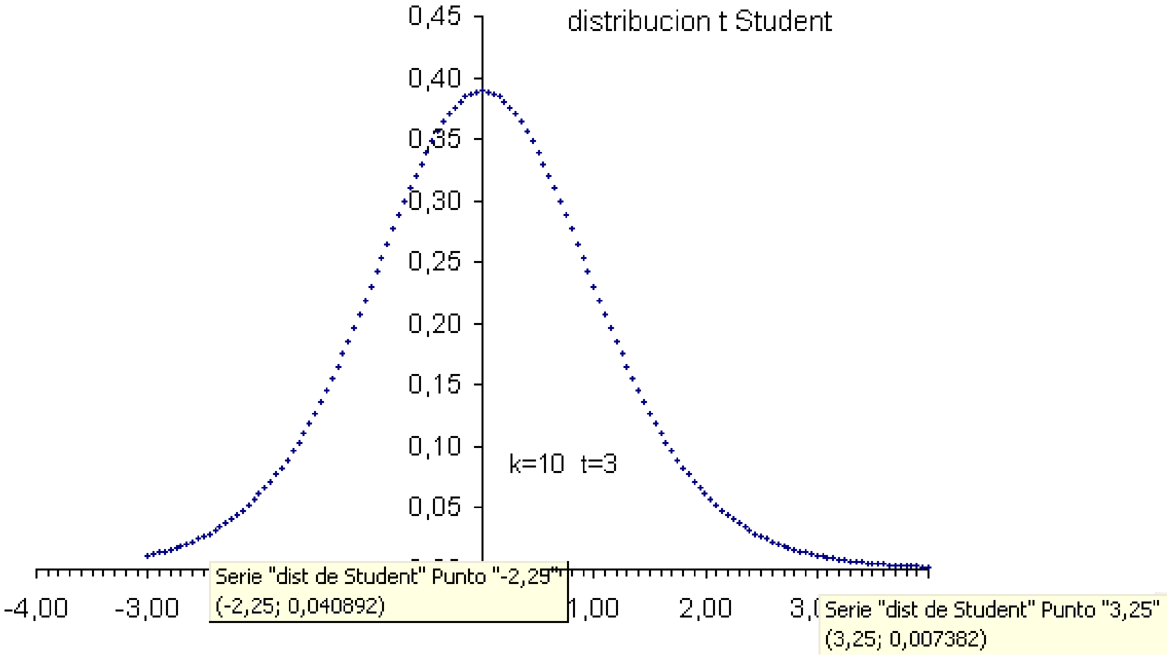
1. Cada curva t tiene forma de campana con centro en 0.
2. Cada curva t, está más dispersa que la curva normal estándar.
3. A medida que k aumenta, la dispersión de la curva t correspondiente disminuye.
4. A medida que $k \rightarrow \infty$, la secuencia de curvas t se aproxima a la curva normal estándar

La distribución de probabilidad de t se publicó por primera vez en 1908 en un artículo de W. S. Gosset. En esa época, Gosset era empleado de una cervecería irlandesa que desaprobaba la publicación de investigaciones de sus empleados. Para evadir esta prohibición, publicó su trabajo en secreto bajo el nombre de "Student". En consecuencia, la distribución t normalmente se llama distribución *t de Student*, o simplemente distribución t.

Ejemplo de calibración

Se desea saber si un instrumento de medición cualquiera está calibrado, desde el punto de vista de la exactitud. Para ello se consigue un valor patrón y se lo mide 10 veces (por ejemplo: una pesa patrón para una balanza, un suero control para un método clínico, etc.). Suponiendo que el resultado de estas mediciones arroja una media de 52,9 y una desviación de 3, usando un patrón de valor 50, se debe determinar si el instrumento está calibrado y la estimación de su error sistemático, si es que se prueba su existencia (no se usan unidades para generalizar este ejemplo).

H_0:μ= 50 el instrumento está calibrado en exactitud
H_1:$\mu \neq 50$ no está calibrado. Hay un error sistemático



Se trata de un ensayo de dos colas donde hay k=10–1=9 grados de libertad. De la Tabla t-Student se obtienen los valores críticos para el 95% de $t_{0,05,9}$=2,262, para el 99% de $t_{0,01,9}$ =3,25 y para un nivel del 99,9% es $t_{0,001,9}$ =4,781. Lo que permite establecer las zonas de aceptación y rechazo:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{(52.9 - 50.0)}{3/\sqrt{10}} = 3$$

Mirando las zonas con los valores críticos, el valor de t cae en la de rechazo para el 95% y no alcanza para las otras. La conclusión es que se ha probado la existencia de un error sistemático con una confianza del 95%.

$$\mu = \bar{x} \pm t\frac{s}{\sqrt{n}} = 52,9 \pm 3\frac{3}{\sqrt{10}} = 52,9 \pm 2,8$$

N	80%	90%	95%	99%	99.9%
1	3,08	6,31	12,7	63,7	637
2	1,89	2,92	4,30	9,92	31,6
3	1,64	2,35	3,18	5,84	12,9
4	1,53	2,13	2,78	4,60	8,60
5	1,48	2,02	2,57	4,03	6,86
6	1,44	1,94	2,45	3,71	5,96
7	1,42	1,90	2,36	3,50	5,40
8	1,40	1,86	2,31	3,36	5,04
9	1,38	1,83	2,26	3,25	4,78
10	1,37	1,81	2,23	3,17	4,59
∞	1,29	1,64	1,96	2,58	3,29

Pude ver en la pagina del laboratorio la hoja de calculo de la Distribución Student, aquí puede observar como esta distribución se aproxima a la distribución de Gauus para medidas mayores a 30.

k es un numero entero positivo k>0

***k*=**	**3**
E[X]=	**0**
Var(X)=	**3,000**
Std. Dev.=	**1,732**

$$h_k(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}}\left(1+\frac{t^2}{k}\right)^{-\frac{(k+1)}{2}}$$

Elavorada por: Lucelly Reyes

Distribución t Student