NOMBRE DEL ALUMNO: YAEL ORLANDO MARTÍNEZ SOLANO

NOMBRE DEL PROFESOR: JORGE SEBASTIÁN DOMÍNGUEZ TORREZ

NOMBRE DEL TRABAJO: TAREA 3

TEMA: LA DISTRIBUCIÓN T DE STUDENT

MATERIA: ESTADÍSTICA INFERENCIAL

GRADO: 4 TRIMESTRE

TEORIAS 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 medida d 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 ala varianza maestral:

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

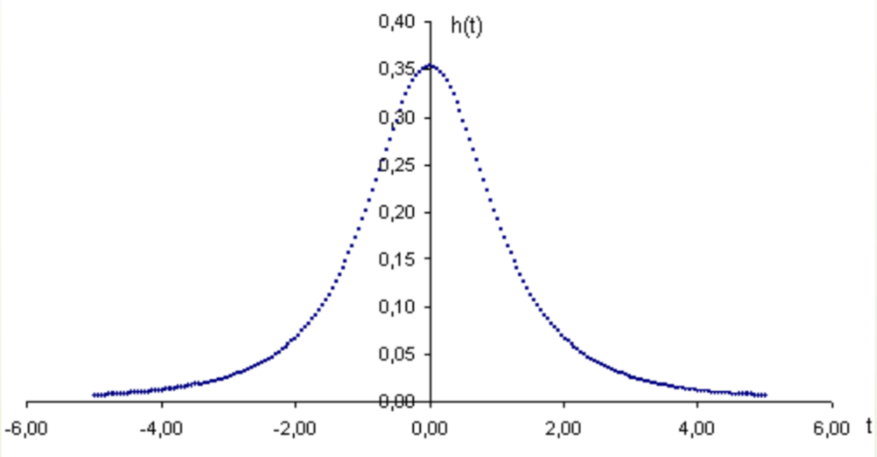
Esta formula esta basada en n-1 grados de libertad. Esta terminología resulta del hecho de que si bien S^2 esta basada en n cantidades $X_1 - \bar{X}, X_2 - \bar{X}, \ldots . X_n - \bar{X}$, estas suman cero, así que especificar los valores de cualquier n-1 de las cantidades determinar el valor restante.

DISTRIBUCION DE PROBABILIDAD T-STUDENT

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

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

- La grafica 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

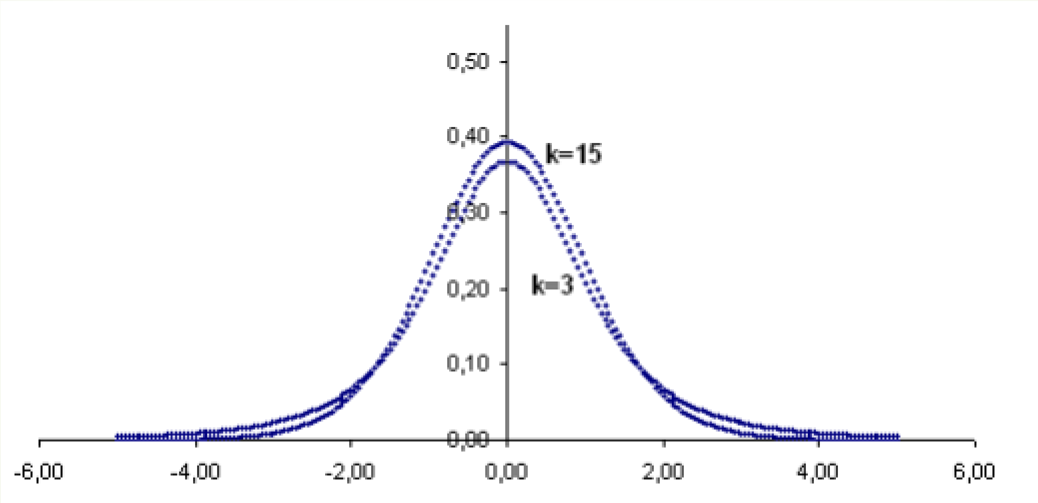
- Su valor medio y varianza es:

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

Si k>3

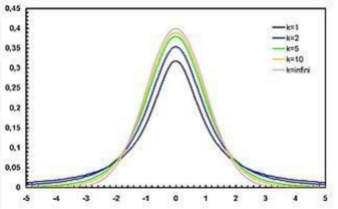
$$Var(T) = \sigma^2 = E((T-\mu)^2) = \int_{-\infty}^{\infty} (t-\mu)^2.h_k(t).dt = \int_{-\infty}^{\infty} t.\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}}\left(1+\frac{t^2}{k}\right)^{-\frac{(k+1)}{2}} dt = ... = \frac{k}{k-2}$$

- La siguiente figura presenta la grafica 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 probabilidad de las colas es mayor que en la distribución normal. A medida que el numero de grados de liberta tiende a infinito, a la forma límite de las distribución t es la distribución normal estándar.

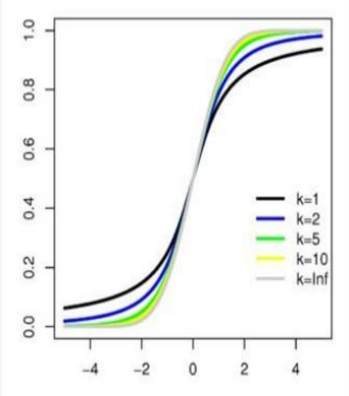


Distribuciones normal y t de student

Distribución t de student



Función de densidad de probabilidad



Función de distribución de probabilidad

Parámetros grados de libertad (real)

Dominio

Función de densidad (pdf)

Media para , indefinida para otros valores

Mediana

Moda

Varianza para , indefinida para otros valores

Coeficiente de simetría para

Curtosis para

Entropía

- : función digamma,
- : función beta

Función generadora de momentos (mgf) (No definida)

[editar datos en Wikidata]

Aparece de manera natural al realizar la prueba t de Student para la determinación de las diferencias entre dos varianzas muéstrales y para la construcción del intervalo de confianza para la diferencia entre las partes de dos poblaciones cuando se desconoce la desviación típica de una población y esta debe ser estimada a partir de los datos de una muestra.

Fue desarrollada por William Sealy Gosset, bajo el seudónimo Student.

CARACTERIZACION

- La distribución t de Student es la distribución de probabilidad del cociente Donde:
- Z es una variable aleatoria distribuida según una normal típica (de media nula y varianza 1).
- V es una variable continua que sigue una distribución χ^2 con grados de libertad.
- Z y V son independientes

Si μ es una constante no nula, el cociente es una variable aleatoria que sigue la distribución t de Student no central con parámetro de no-centralidad .

- Aparición y especificaciones de la distribución t de Student

Supongamos que X_1,..., X_n son variables aleatorias independientes distribuidas normalmente, con media μ y varianza σ^2. Sea la media muestra, Entonces sigue una distribución normal de media 0 y varianza 1.

Sin embargo, dado que la desviación estándar no siempre es conocida de antemano, Gosset estudió un cociente relacionado, es la cuasi varianza maestral y demostró que la función de densidad de T es donde es igual a n − 1. La distribución de T se llama ahora la distribución-t de Student.

El parámetro representa el número de grados de libertad. La distribución depende de , pero no de o , lo cual es muy importante en la práctica. Intervalos de confianza derivados de la distribución t de Student. El procedimiento para el cálculo del intervalo de confianza basado en la t de Student consiste en estimar la desviación típica de los datos S y calcular el error estándar de la media: ,siendo entonces el intervalo de confianza para la media: Es este resultado el que se utiliza en el test de Student: puesto que la diferencia de las medias de muestras de dos distribuciones normales se distribuye también normalmente, la distribución t puede usarse para examinar si esa diferencia puede razonablemente suponerse igual a cero.

Para efectos prácticos el valor esperado y la varianza son: y para Historia La distribución de Student fue descrita en el año 1908 por William Sealy Gosset. Gosset trabajaba en una fábrica de cerveza, Guinness, que prohibía a sus empleados la publicación de artículos científicos debido a una difusión previa de secretos industriales. De ahí que Gosset publicase sus resultados bajo el seudónimo de Student.

DISTRIBUCION “T” DE STUDENT

- Descrita por William S. Gosset en 1908. Publicaba bajo el pseudónimo de “Student” mientras trabajaba para la cervecería Guinnes en Irlanda. Está diseñada para probar hipótesis en estudios con muestras pequeñas (menores de 30) La fórmula general para la T de Student es la siguiente:

$$t = \frac{X - \mu}{s/\sqrt{n}}$$

En donde el numerador representa la diferencia a probar y el denominador la desviación estándar de la diferencia llamado también Error Estándar. En esta fórmula t representa al valor estadístico que estamos buscando X barra es el promedio de la variable analizada de la muestra, y miu es el promedio poblacional de la variable a estudiar. En el denominador tenemos a s como representativo de la desviación estándar de la muestra y n el tamaño de ésta.

Grados de libertad: El número de grados de libertad es igual al tamaño de la muestra (número de observaciones independientes) menos 1.

gl = df = (n – 1)

- Si pudiera expresar en un cierto número de pasos para resolver un problema de t de student tendría que declarar los siguientes:

- Paso 1. Plantear la hipótesis nula (H_0) y la hipótesis alternativa (H_1). La hipótesis alternativa plantea matemáticamente lo que queremos demostrar, en tanto que la hipótesis nula plantea exactamente lo contrario.
- Paso 2. Determinar el nivel de significancia (rango de aceptación de la hipótesis alternativa), a. Se considera un nivel alfa de: 0.05 para proyectos de investigación; 0.01 para aseguramiento de la calidad; y 0.10 para estudios o encuestas de mercadotecnia.
- Paso 3. Evidencia muestral, se calcula la media y la desviación estándar a partir de la muestra.
- Paso 4. Se aplica la distribución T de Student para calcular la probabilidad de error por medio de la fórmula general presentada al principio y se contrasta con el valor T obtenido de la tabla correspondiente.
- Paso 5. En base a la evidencia disponible se acepta o se rechaza la hipótesis alternativa. Si la probabilidad de error (p) es mayor que el nivel de significancia se rechaza la hipótesis alternativa. Si la probabilidad de error (p) es menor que el nivel de significancia se acepta la hipótesis alternativa.

PRUEBA T DE STUDENT

- Con esta prueba se pretende averiguar si dos muestras que tienen medias iguales, provienen de la misma población.

Hipótesis nula "H_0" → $\mu_1 = \mu_2$;

Hipótesis alternativa "H_1" → $\mu_1 \neq \mu_2$

La prueba permite comparar la media con su valor verdadero o bien las medias de dos poblaciones. Se basa en los límites de confianza "LC" para el promedio x de n mediciones repetidas (Ec. 2.1). A partir de dicha ecuación tenemos:

$\mu = x \pm t(s/\sqrt{n})$ (Ec. 2.1) → $x - \mu = \pm t\, s/\sqrt{n}$ (Ec. 2.2)

$s/\sqrt{n}$: error estándar "EE" o desviación estándar "DE" de la distribución muestral de medias. Como las medias son $\sqrt{n}$ veces más probables que los resultados aislados, la DE de las medias es $\sqrt{n}$ veces menor que la DE de resultados aislados, siendo n el número de determinaciones con las que se calcula la media.

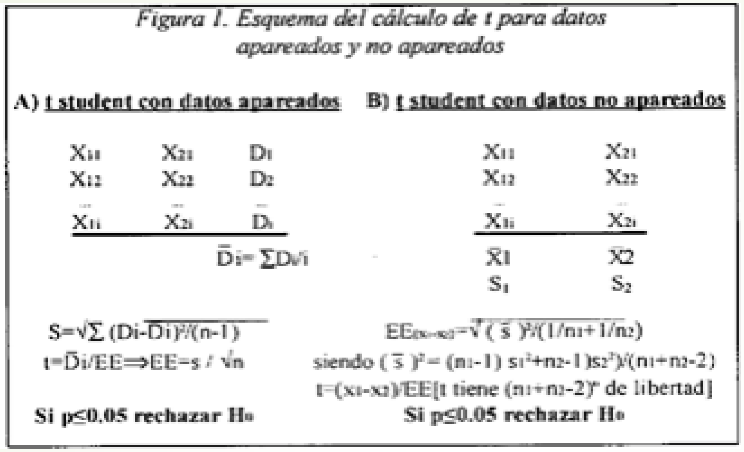
t: "t de student" (tabla 2). Es un parámetro tabulado que depende de los grados de libertad de la muestra (n-1) "gl" y del intervalo de confianza que se quiera (generalmente 95%).

gl	1	2	3	4	5	6	7	8	9
t	12.71	4.3	3.18	2.78	2.57	2.45	2.36	2.31	2.26
	(6.31)	(2.92)	(2.35)	(2.13)	(2.02)	(1.94)	(1.89)	(1.86)	(1.83)
gl	**10**	**12**	**14**	**16**	**18**	**20**	**30**	**50**	**α**
T	2.23	2.18	2.14	2.12	2.1	2.09	2.04	2.01	1.96
	(1.81)	(1.78)	(1.76)	(1.75)	(1.73)	(1.72)	(1.70)	(1.68)	(1.64)

Nota: *los valores críticos de t son estimados para una prueba de 2 colas. Para una prueba de una cola se toma el valor que corresponde a p=0.1, es decir, el doble del valor de p deseado (0.05).*

Si $x - \mu$ obtenida en la muestra a comparar es menor que la calculada para un cierto nivel de probabilidad, no se rechaza la hipótesis nula de que x y μ sean iguales; es decir, sus diferencias son debidas a errores aleatorios y no existe un error sistemático significativo.

- Para comparar 2 medias experimentales el proceso es semejante. Se ha de tener en cuenta si los datos de las 2 muestras están apareados o no (figura 1):



- Datos apareados: tienen la ventaja de permitir trabajar simplificando a una sola muestra (cuyos valores corresponden a la diferencia "D_i" entre cada par de datos apareados). Sustituimos x - µ (Ec. 2.2) por Di - 0 porque el valor real de las diferencias, suponiendo que las dos muestras tienen la misma media, es 0. La DE se calcula con la muestra de diferencias.

- Datos no apareados: como no se puede simplificar a una sola muestra, se ha de introducir el concepto de desviación estándar ponderada "sp" (Ec. 2.3). En la ecuación 2.2 se sustituye s por S_p y x - µ por X_1 - X_2 y el tamaño de muestra "n" se sustituye por N ponderado "(N_1 + N_2)/ N1N2".

- S_p=√[S(X_1 - X_1)2+S(X_2 - X_2)2+ ...]/(N_1+ N_2+ ... - N_s) (Ec. 2.3)

- N_1 , N_2, ...: el tamaño de las muestras.

- N_s: número de muestras.

- (n_1+n_2+...-N_s): número de grados de libertad.

AQUÍ CONCLUYO MI TRABAJO, MUCHAS GRACIAS POR SU ATENCION PRESTADA, LE DESEO UN EXELENTE DIA, TARDE O NOCHE, DEPENDE EN QUE MOMENTO REVISO MI TRABJO ESPERO QUE HAYA SIDO DE SU AGRADO, SALUDOS LICENCIADO.