Distribuciones Continuas de Probabilidad:
Normal y "t" de Student

Penna – Cobos – Vázquez Ferrero – Ulagnero

DISTRIBUCIONES CONTINUAS DE PROBABILIDAD

INTRODUCCIÓN

Las distribuciones de probabilidad son modelos matemáticos teóricos que asignan, a priori, la frecuencia relativa esperada a cada categoría o intervalo de clase factible para una variable aleatoria.

- Cualquiera podrá ser la variable aleatoria a distribuir, siempre y cuando cumpla con un procedimiento estocástico -o sea, de asignación al azar- en la determinación de los valores que corresponderán a cada UA: esto es lo que define como aleatoria a una variable.
- Luego, entonces, se podrá con tales distribuciones conocer la probabilidad de ocurrencia de cada valor de esa variable. El modelo matemático encuentra, en esto, su principal utilidad predictiva.
- Como venimos sabiendo, la variable será considerada continua cuando sea plausible que pueda tomar los infinitos valores no numerables -lo que es equivalente a decir que es posible asignar cualquiera de los valores de cualquier fraccionamiento- de un intervalo.

DISTRIBUCIÓN NORMAL

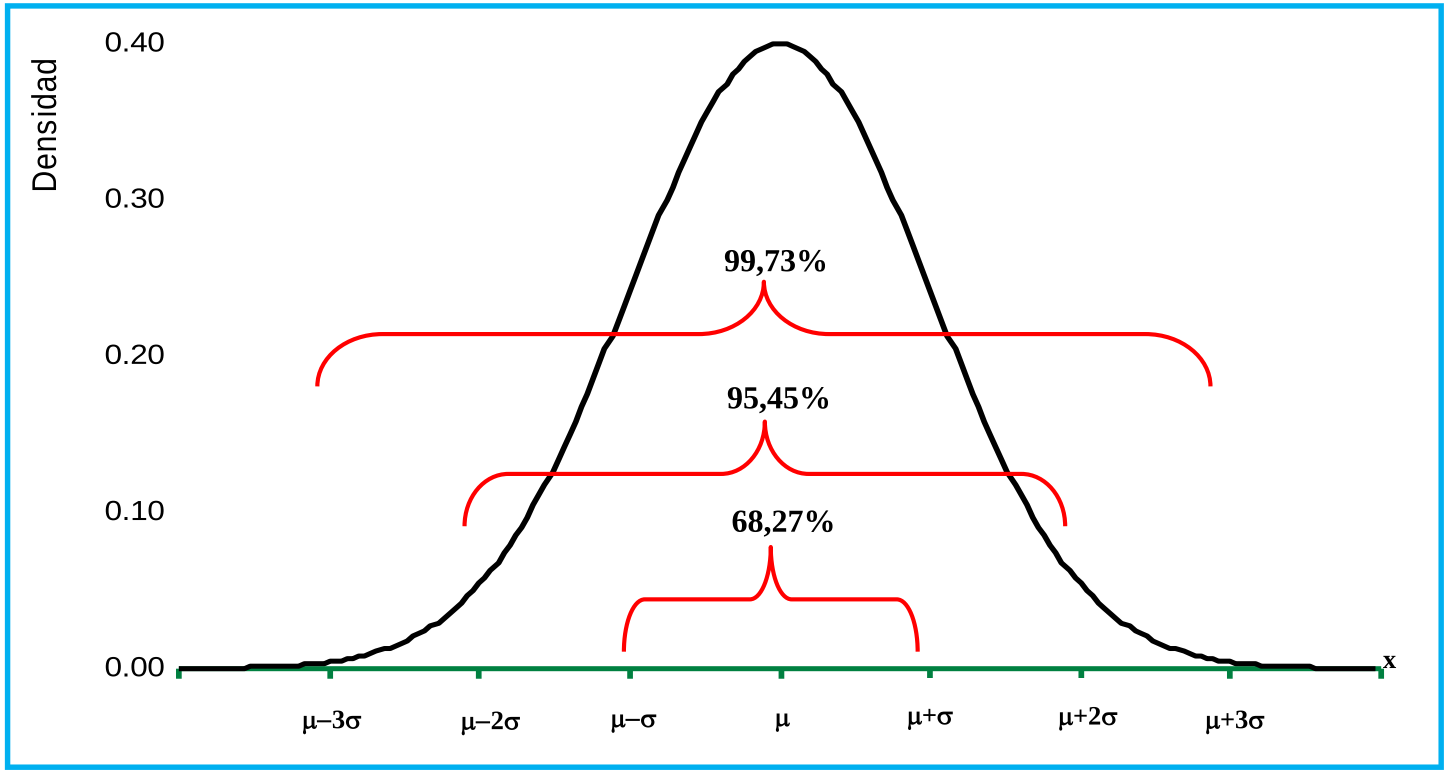
Esta función tiene gran importancia teórica tanto en problemas de tipo biológico como sociales, físicos, químicos, etc. La distribución normal (también conocida como distribución de Laplace-Gauss, campana de Gauss o, simplemente, Gaussiana), tiene su origen entre los siglos XVII y XVIII como primera aproximación al comportamiento de los errores pues, investigadores como De Moivre, Laplace y Gauss, encontraron cierta "regularidad" en los mismos. Esto motivó que, en sus comienzos, a dicha curva se conociese como "curva normal de errores".

Como dijimos, a esta distribución se la utiliza en presencia de variables continuas tales como peso, altura, edad, CI, IMC, etc., y su expresión matemática, a partir de una variable continua X, se define como:

$$\boxed{\mathbf{f(x)} = \frac{\mathbf{1}}{\boldsymbol{\sigma}\sqrt{\mathbf{2}\boldsymbol{\pi}}}\mathbf{e}^{-\frac{\mathbf{1}}{\mathbf{2}}\left(\frac{\mathbf{x}-\boldsymbol{\mu}}{\boldsymbol{\sigma}}\right)^{\mathbf{2}}} \quad \begin{cases} x \in (-\infty, +\infty) \\ \mu \in (-\infty, +\infty) \\ \sigma > 0 \end{cases}}$$

DISTRIBUCIÓN NORMAL (CONT.)

Siendo su representación gráfica, la siguiente:



Esta curva depende, básicamente, de los parámetros $\boldsymbol{\mu}$ (media aritmética o esperanza matemática) y $\boldsymbol{\sigma^2}$ (varianza). Por este motivo, dada una variable aleatoria X, distribuida normalmente y caracterizada por su media, μ, y su desviación estándar, σ, se la puede expresar como: $\mathbf{X} \backsim \boldsymbol{N}(\boldsymbol{\mu};\ \boldsymbol{\sigma})$

DISTRIBUCIÓN NORMAL (CONT.)

Donde, algunas de sus características son las siguientes:

- Es una curva unimodal.
- El valor de la moda coincide con los valores de la media y la mediana.
- Presenta una forma de campana.
- Es unitaria.
- Es simétrica respecto a su media, μ.
- Es asintótica respecto al eje horizontal.
- Los puntos de inflexión se encuentran en $\mu - \sigma$ y $\mu + \sigma$.
- El área, bajo la curva, comprendida en el intervalo $(\mu - \sigma; \mu + \sigma)$ es igual a 0,6827 (68,27%); entre $(\mu - 2\sigma; \mu + 2\sigma)$ es igual a 0,9545 (95,45%) y entre $(\mu - 3\sigma; \mu + 3\sigma)$ es igual a 0,9973 (99,73%), como se vio en el gráfico anterior.

DISTRIBUCIÓN NORMAL (CONT.)

ESTANDARIZACIÓN

Por lo visto anteriormente se deduce que no existe una única distribución normal, sino una familia de distribuciones normales, diferenciadas por los valores de su media (μ) y su varianza (σ^2). Es por eso que si necesitáramos determinar una "porción de área" bajo la curva, tendríamos que realizar engorrosos cálculos matemáticos para cada curva normal. Mediante la siguiente transformación, se lleva la variable original de puntajes "x" a puntajes "z":

$$\boxed{\mathbf{X} \backsim N(\mu;\ \sigma) \Rightarrow \mathbf{Z} = \frac{\mathbf{x} - \mu}{\sigma} \backsim N(0;\ 1)}$$

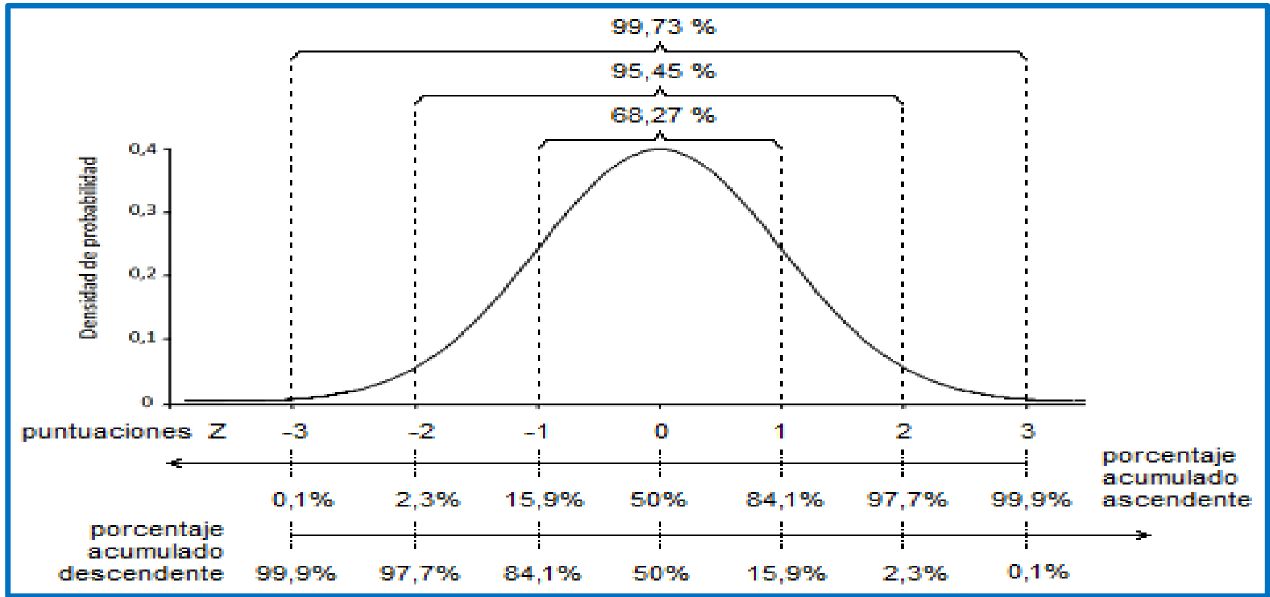
Esta propiedad resulta interesante en la práctica, ya que para una distribución $N(0;\ 1)$ existen tablas a partir de las que se pueden obtener, de modo sencillo, la probabilidad de observar un área menor o mayor a un cierto valor z, y que permitirán resolver problemas de probabilidad sobre el comportamiento de variables de las que se sabe -o se asume- que siguen una distribución aproximadamente normal.

Distribución Normal (cont.)

Como consecuencia de lo expresado, podemos decir que "una puntuación z nos indica la dirección y grado en que un valor individual obtenido se aleja de la media, en una escala de unidades de desviación estándar".

Frente a la transformación realizada, según corresponda, la curva normal teórica se reduce a la siguiente expresión:

$$\boxed{\mathbf{f(z)} = \frac{\mathbf{1}}{\sqrt{\mathbf{2\pi}}}\mathbf{e}^{-\frac{\mathbf{z^2}}{\mathbf{2}}} \quad \text{para } z \in (-\infty, +\infty)}$$



DISTRIBUCIÓN NORMAL (EJEMPLO)

Las puntuaciones en la Escala de Inteligencia para adultos siguen, en una determinada población, una Distribución Normal de media 100 y desviación estándar 16. Si extraemos de dicha población un sujeto de manera aleatoria, cuál es la probabilidad que tenga un puntaje:

a) ¿Inferior a 94,08? **b)** ¿Entre 96 y 104? **c)** ¿Superior a 108?

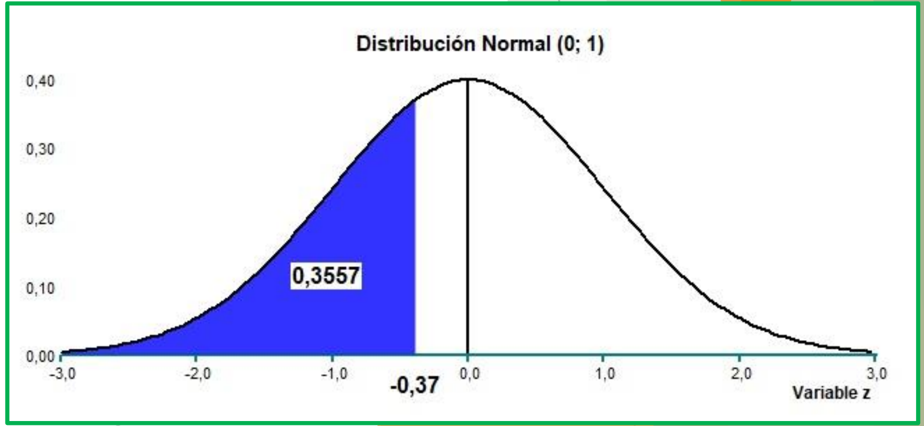
Solución: de acuerdo a los datos del problema, podemos ver que:

$$\text{X: “escala de inteligencia para adultos”} \Rightarrow X \sim N(100; 16)$$

a) $P(X < 94{,}08) = P\left(Z < \frac{94{,}08 - 100}{16}\right) = P(Z < -0{,}37) = P(Z > 0{,}37) = \Phi(0{,}37) = \mathbf{0{,}3557}$

TABLA 1

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660
1,0	0,1587	0,1562	0,1539	0,1515	0,1492			
1,1	0,1357	0,1335	0,1314	[illegible]				



InfoStat

DISTRIBUCIÓN NORMAL (EJEMPLO)

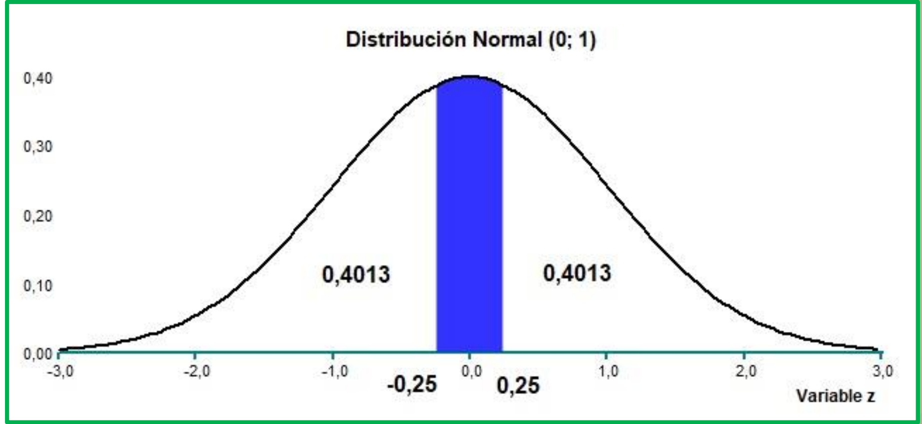
b) $P(96 < X < 104) = P\left(\frac{96-100}{16} < Z < \frac{104-100}{16}\right) = P(-0{,}25 < Z < 0{,}25) =$

$= 1 - [P(Z > 0{,}25) + P(Z > 0{,}25)] = 1 - 2 \times P(Z > 0{,}25) = 1 - 2 \times \Phi(0{,}25) =$

$= 1 - 2 \times 0{,}4013 = 1 - 0{,}8026 = \mathbf{0,1974}$

TABLA 1

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660
1,0	0,1587	0,1562	0,1539	0,1515	0,1492			
1,1	0,1357	0,1335	0,1314	[illegible]				

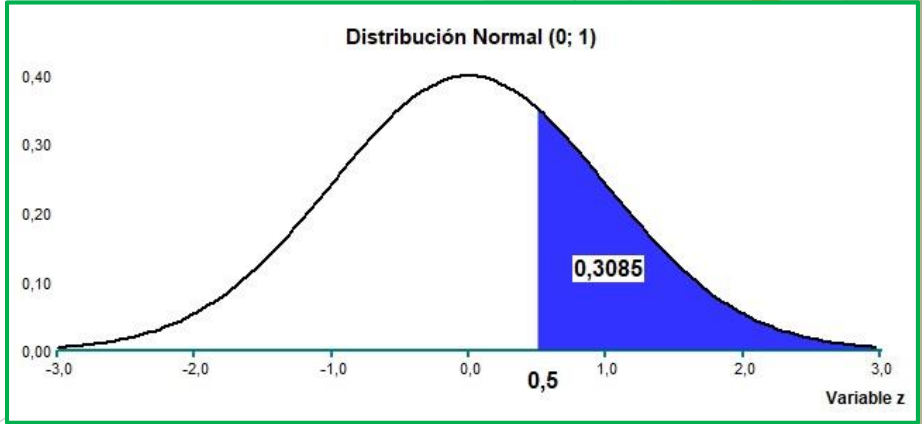


c) $P(X > 108) = P\left(Z > \frac{108-100}{16}\right) = P(Z > 0{,}5) = \Phi(0{,}5) = \mathbf{0,3085}$

TABLA 1

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660
1,0	0,1587	0,1562	0,1539	0,1515	0,1492			
1,1	0,1357	0,1335	0,1314	[illegible]				

InfoStat



Distribución "t" de Student

Como se mencionó, una de las principales aplicaciones del modelo de distribución normal de probabilidades en el campo de las investigaciones, es el estudio de distribuciones de *medias* para inferir -valga la redundancia- *parámetros poblacionales*, mediante *muestras*.

Dedicaremos las siguientes Unidades de la materia a su estudio, pero presentamos en esta ocasión el modelo de distribución de probabilidades ideado por W. Gosset (1876-1937), discípulo de Fisher, razón esta de su elección del seudónimo "*Student*" para la publicación de sus investigaciones de manera anónima cumpliendo sus compromisos contractuales con la empresa fabricante de cerveza Guinnes, para la cual trabajaba.

Podemos entender la distribución "t" de Student como un desarrollo de la tecnología estadística que aumenta el nivel de precisión en la estimación de probabilidades a partir de <u>muestras pequeñas</u>. No representa, entonces, una forma en que se distribuyan mediciones reales; ni se diferencia de una distribución normal estándar en el caso de muestras muy grandes.

DISTRIBUCIÓN “t” DE STUDENT (CONT.)

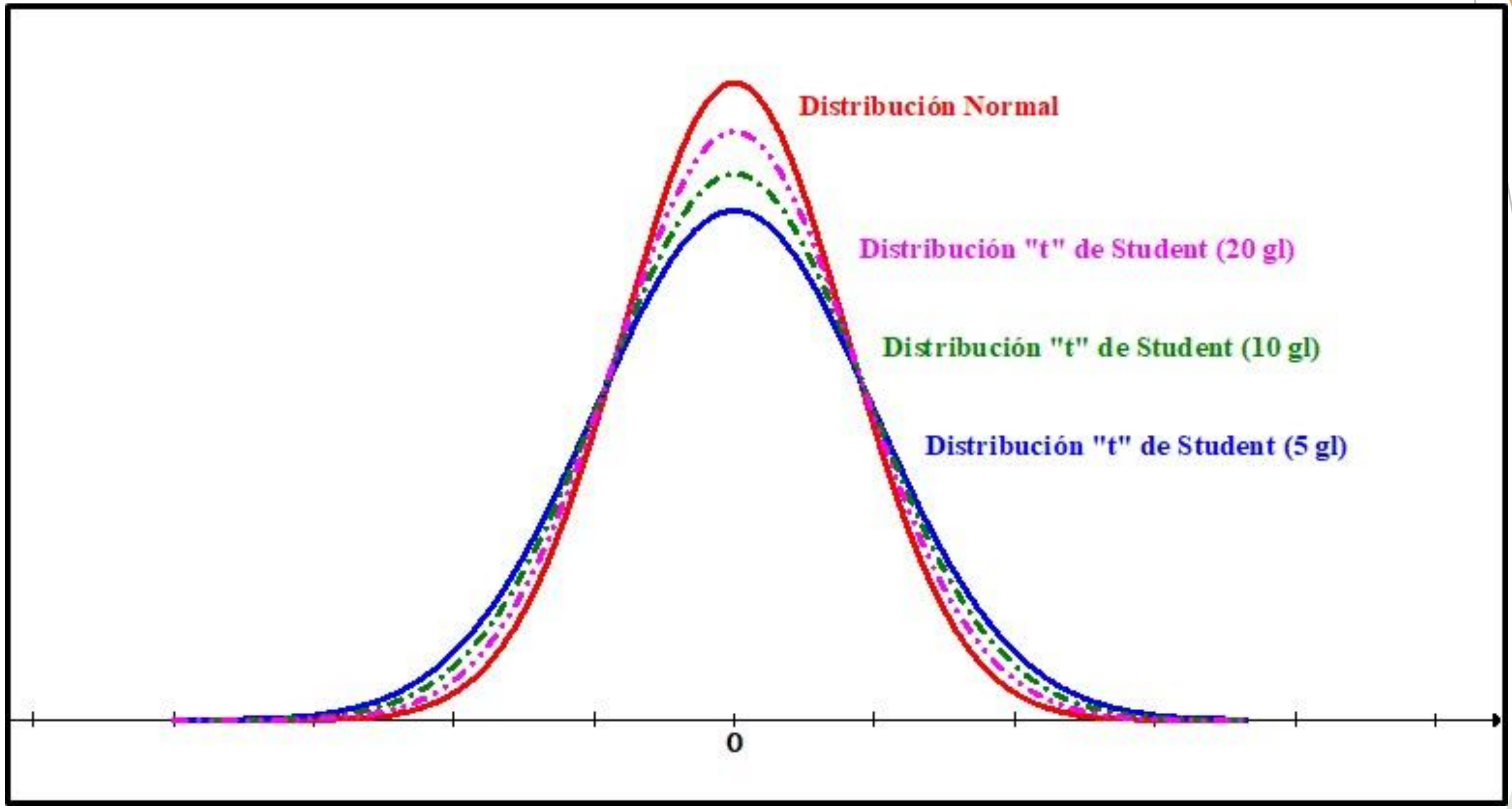
La curva de su función de densidad queda definida por sus grados de libertad (*gl*), por lo que habrá infinitas distribuciones “t” dependiendo de tal valor; por otro lado, comparte las mismas propiedades geométricas de la curva normal -incluso la forma acampanada- pero sus puntos de inflexión se encuentran más alejados respecto de su media, razón por la cual su altura es más baja y sus colas más anchas.

Sin embargo, a medida en que se aumenten sus grados de libertad, tenderá a cobrar la forma de la distribución normal estándar (igualándola al trabajar con muestras grandes).

Por lo general, al trabajar con muestras, los datos estarán menos dispersos que en el total de la población y la varianza calculada en aquéllas, entonces, será menor. Para corregir tal sesgo, las fórmulas que la estiman partiendo de muestras procuran un resultado más fidedigno -más grande- dividiendo los valores por un número menor. El denominador requerido corresponderá con un valor mínimamente inferior al tamaño total de la muestra, y se obtendrá restando una unidad de análisis al tamaño total de dicha muestra: $n - 1$.

DISTRIBUCIÓN "t" DE STUDENT (CONT.)

Gráficamente:



Según lo dicho, la distribución "t" de Student se aplica en las llamadas *Pruebas T* también para el cálculo de probabilidades pero preferentemente cuando se cuenta con muestras pequeñas (de 30 UA o menos); y para la estimación de la varianza poblacional a partir de los valores muestrales (es decir, cuando desconocemos tal varianza).

DISTRIBUCIÓN "t" DE STUDENT (EJEMPLO)

Los siguientes ejercicio, nos sirven para conocer la utilización de la tabla "t" de Student.

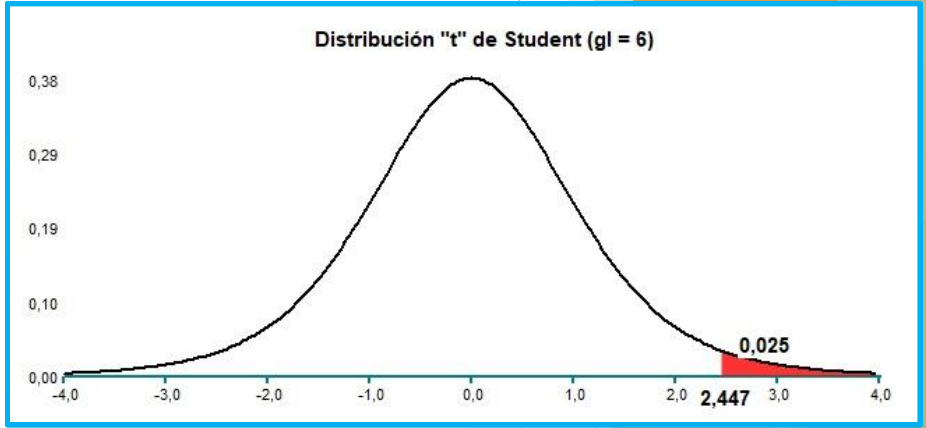
a) $P(t_6 \geq t_0) = 0{,}025$; **b)** $P(t_{11} < t_0) = 0{,}90$; **c)** $P(t_7 > 2{,}873) = \alpha$; **d)** $P(t_4 \leq 1{,}272) = \alpha$

Solución:

a) $P(t_6 \geq t_0) = 0{,}025 \Rightarrow t_0 = \mathbf{2,447}$

TABLA 2

gl \ α	0,25	0,20	0,15	0,10	0,05	0,025	0,01	0,005	0,0005
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,656	636,578
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,600
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,718	0,906	1,134	1,440	1,943	$\boxed{2,447}$	3,143	3,707	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624	[illegible]	
15	0,691	0,866	1,074	1,341	1,753	[illegible]			
16	0,690	0,865	1,071	1,337					



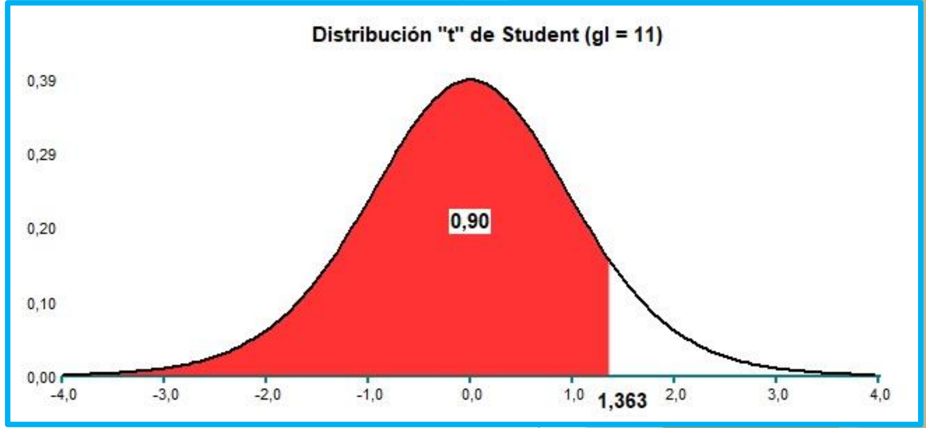
InfoStat

Distribución "t" de Student (ejemplo)

b) $P(t_{11} < t_0) = 0{,}90 \Rightarrow P(t_{11} > t_0) = 0{,}10 \Rightarrow t_0 = \mathbf{1{,}363}$

TABLA 2

gl \ α	0,25	0,20	0,15	0,10	0,05	0,025	0,01	0,005	0,0005
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,656	636,578
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,600
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	0,876	1,088	$\boxed{1,363}$	1,796	2,201	2,718	3,106	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624		
15	0,691	0,866	1,074	1,341	1,753				
16	0,690	0,865	1,071	1,337					

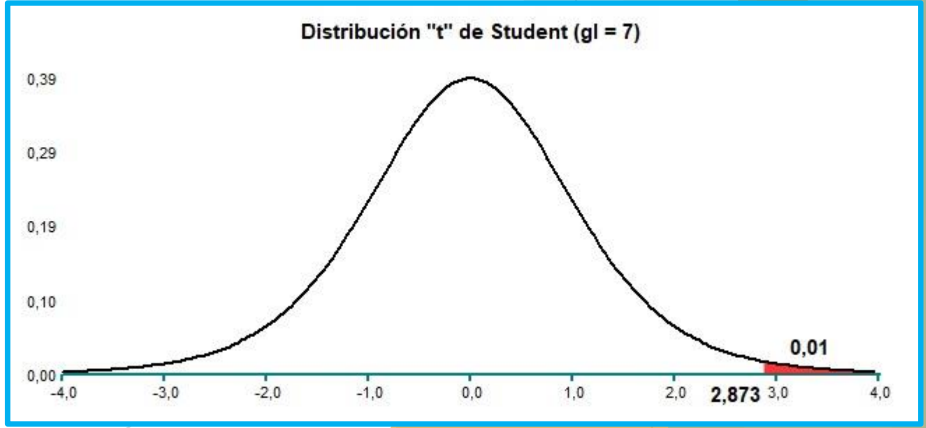


InfoStat

c) $P(t_7 > 2{,}873) = \alpha \Rightarrow \alpha \cong \mathbf{0{,}01}$

TABLA 2

gl \ α	0,25	0,20	0,15	0,10	0,05	0,025	0,01	0,005	0,0005
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,656	636,578
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,600
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	$\boxed{2,998}$	3,499	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624		
15	0,691	0,866	1,074	1,341	1,753				
16	0,690	0,865	1,071	1,337					

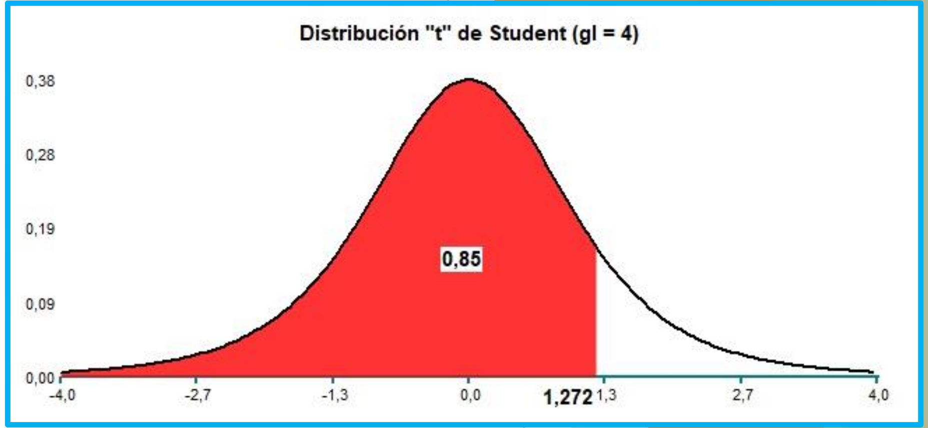


DISTRIBUCIÓN “t” DE STUDENT (EJEMPLO)

d) $P(t_4 \le 1{,}272) = \alpha \Rightarrow P(t_4 \le 1{,}272) = 1 - P(t_4 > 1{,}272) \cong 1 - 0{,}15 = \mathbf{0{,}85}$

TABLA 2

gl \ α	0,25	0,20	0,15	0,10	0,05	0,025	0,01	0,005	0,0005
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,656	636,578
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,600
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,741	0,941	$\boxed{1,190}$	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624	[illegible]	
15	0,691	0,866	1,074	1,341	1,753	2,131			
16	0,690	0,865	1,071	1,337					



InfoStat

Referencias Bibliográficas

- Bologna, E. (2011). *Estadística para Psicología y Educación*. Córdoba: Brujas.
- Glass, V. & Stanley, J.C. (1996). *Métodos Estadísticos aplicados a las Ciencias Sociales*. México: Prentice-Hall Hispanoamericana, S.A.
- Gorgas García, J., Cardiel López, N. & Zamorano Calvo, J. (2009). *Estadística Básica para Estudiantes de Ciencias*. Madrid: Departamento de Astrofísica y Ciencias de la Atmósfera. Facultad de Ciencias Físicas. Universidad Complutense de Madrid.
- Hernández Sampieri, R., Fernández-Collado, C., Baptista Lucio, P. (2014). *Metodología de la investigación* (6ª ed.). México: McGraw-Hill Interamericana.
- Penna, F.O., Esteva, G.C., Cobos, O.H. & Ulagnero, C.A. (2018). *Fórmulas y Tablas III (para cursos de Estadística básica)* (2ª ed.). San Luis: Nueva Editorial Universitaria.
- Triola M. (2018). *Estadística* (12ª ed.). México: Pearson Educación.