Distribución t de Student empleando Excel y Winstats

1) Introducción

Al comenzar el siglo XX, un especialista en Estadística de la Guinness Breweries en Irlanda llamado William S. Gosset deseaba hacer inferencias acerca de la media cuando la σ fuera desconocida. Como a los empleados de Guinness no se les permitía publicar el trabajo de investigación bajo sus propios nombres, Gosset adoptó el seudónimo de "Student". La distribución que desarrolló se conoce como la distribución t de Student.

Si la variable aleatoria X se distribuye normalmente, entonces el siguiente estadístico tiene una distribución t con n - 1 grados de libertad.

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

Esta expresión tiene la misma forma que el estadístico Z en la ecuación para la distribución muestral de la media con la excepción de que S se usa para estimar la σ desconocida.

Entre las principales propiedades de la distribución t se tiene:

En apariencia, la distribución t es muy similar a la distribución normal estandarizada. Ambas distribuciones tienen forma de campana. Sin embargo, la distribución t tiene mayor área en los extremos y menor en el centro, a diferencia de la distribución normal. Puesto que el valor de σ es desconocido, y se emplea S para estimarlo, los valores t son más variables que los valores Z.
Los grados de libertad n - 1 están directamente relacionados con el tamaño de la muestra n. A medida que el tamaño de la muestra y los grados de libertad se incrementan, S se vuelve una mejor estimación de σ y la distribución t gradualmente se acerca a la distribución normal estandarizada hasta que ambas son virtualmente idénticas. Con una muestra de 120 o más, S estima σ con la suficiente precisión como para que haya poca diferencia entre las distribuciones t y Z. Por esta razón, la mayoría de los especialistas en estadística usan Z en lugar de t cuando el tamaño de la muestra es igual o mayor de 30.

Como se estableció anteriormente, la distribución t supone que la variable aleatoria X se distribuye normalmente. En la práctica, sin embargo, mientras el tamaño de la muestra sea lo suficientemente grande y la población no sea muy sesgada, la distribución t servirá para estimar la media poblacional cuando σ sea desconocida.

Los grados de libertad de esta distribución se calculan con la siguiente fórmula

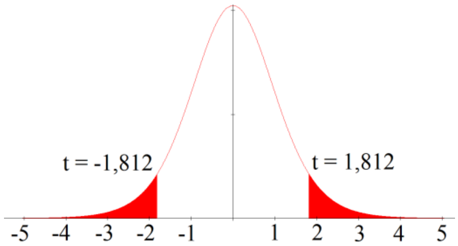
$$n - 1$$

Donde n = tamaño de la muestra

Ejemplo: Imagínese una clase con 40 sillas vacías, cada uno elige un asiento de los que están vacíos. Naturalmente el primer alumno podrá elegir de entre 40 sillas, el segundo de entre 39, y así el número irá disminuyendo hasta que llegue el último alumno. En este punto no hay otra elección (grado de libertad) y aquel último estudiante simplemente se sentará en la silla que queda. De este modo, los 40 alumnos tienen 39 o n-1 grados de libertad.

Para leer en la tabla de la distribución t se procede de la siguiente manera:

TABLA
DISTRIBUCIÓN t DE STUDENT



Ejemplos:
Para n-1 = 10 grados de libertad
$\boldsymbol{P(t > 1,812) = 0,05}$
$\boldsymbol{P(t < -1,812) = 0,05}$

α / n-1	0,25	0,2	0,15	0,1	**0,05**	0,025	0,01	0,005	0,0005
1	1,0000	1,3764	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567	636,6192
2	0,8165	1,0607	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248	31,5991
3	0,7649	0,9785	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409	12,9240
4	0,7407	0,9410	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041	8,6103
5	0,7267	0,9195	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321	6,8688
6	0,7176	0,9057	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074	5,9588
7	0,7111	0,8960	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995	5,4079
8	0,7064	0,8889	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554	5,0413
9	0,7027	0,8834	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498	4,7809
10	0,6998	0,8791	1,0931	1,3722	**1,8125**	2,2281	2,7638	3,1693	4,5869
11	0,6974	0,8755	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058	4,4370

Usted encontrará los valores críticos de t para los grados de libertad adecuados en la tabla para la distribución t. Las columnas de la tabla representan el área de la cola superior de la distribución t. Cada fila representa el valor t determinado para cada grado de libertad específico. Por ejemplo, con 10 grados de libertad, si se quiere un nivel de confianza del 90%, se encuentra el valor t apropiado como se muestra en la tabla. El nivel de confianza del 90% significa que el 5% de los valores (un área de 0,05) se encuentran en cada extremo de la distribución. Buscando en la columna para un área de la cola superior y en la fila correspondiente a 10 grados de libertad, se obtiene un valor crítico para t de 1.812. Puesto que t es una distribución simétrica con una media 0, si el valor de la cola superior es +1.812, el valor para el área de la cola inferior (0,05 inferior) sería -1.812. Un valor t de -1.812 significa que la probabilidad de que t sea menor a -1.812, es 0,05, o 5% (vea la figura).

2) Ejemplos ilustrativos:

2.1) Determinar el valor crítico de t con lectura en la tabla, Excel y Winstats en cada una de las siguientes condiciones para $1-\alpha = 0{,}95\,;n = 13$

Solución:

Con lectura en la tabla

$$Si\ 1-\alpha = 0{,}95 \Rightarrow \alpha = 1-0{,}95 = 0{,}05$$

Para leer en la tabla se necesita calcular el área de una cola, la cual es:

$$\frac{\alpha}{2} = \frac{0{,}05}{2} = 0{,}025$$

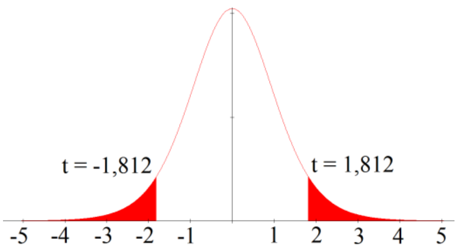
O también el área de una cola se calcula de la siguiente manera:

$$\frac{\alpha}{2} = \frac{1-(1-\alpha)}{2} \Rightarrow \frac{\alpha}{2} = \frac{1-0{,}95}{2} = 0{,}025$$

Calculando los grados de libertad se tiene:

$n - 1 = 13 - 1 = 12$

TABLA
DISTRIBUCIÓN t DE STUDENT



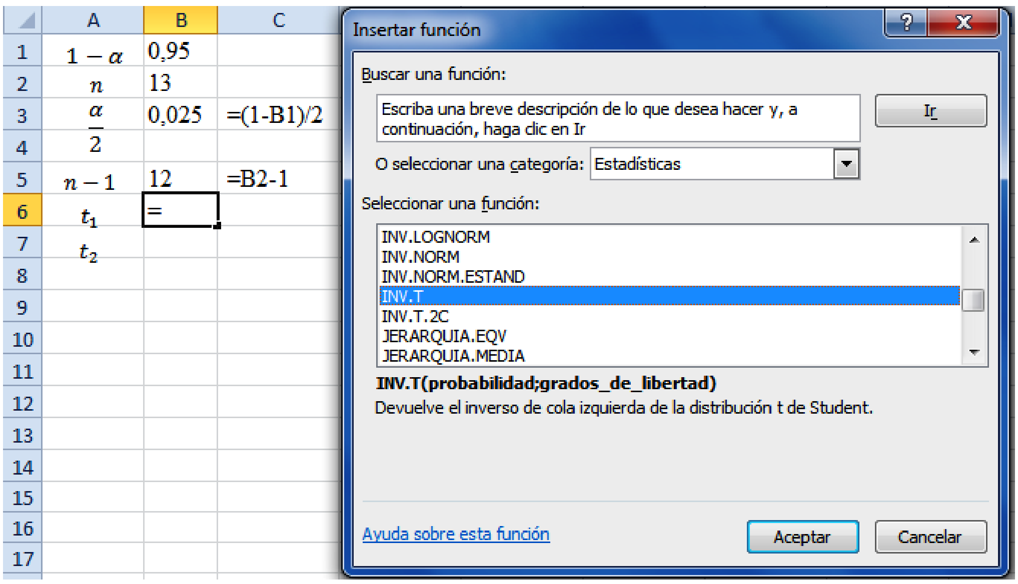
Ejemplos:
Para n-1 = 10 grados de libertad
$\boldsymbol{P(t > 1{,}812) = 0{,}05}$
$\boldsymbol{P(t < -1{,}812) = 0{,}05}$

n-1 \ α	0,25	0,2	0,15	0,1	0,05	**0,025**	0,01	0,005	0,0005
1	1,0000	1,3764	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567	636,6192
2	0,8165	1,0607	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248	31,5991
3	0,7649	0,9785	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409	12,9240
4	0,7407	0,9410	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041	8,6103
5	0,7267	0,9195	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321	6,8688
6	0,7176	0,9057	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074	5,9588
7	0,7111	0,8960	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995	5,4079
8	0,7064	0,8889	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554	5,0413
9	0,7027	0,8834	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498	4,7809
10	0,6998	0,8791	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693	4,5869
11	0,6974	0,8755	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058	4,4370
12	0,6955	0,8726	1,0832	1,3562	1,7823	**2,1788**	2,6810	3,0545	4,3178
13	0,6938	0,8702	1,0795	1,3502	1,7709	2,1604	2,6503	3,0123	4,2208

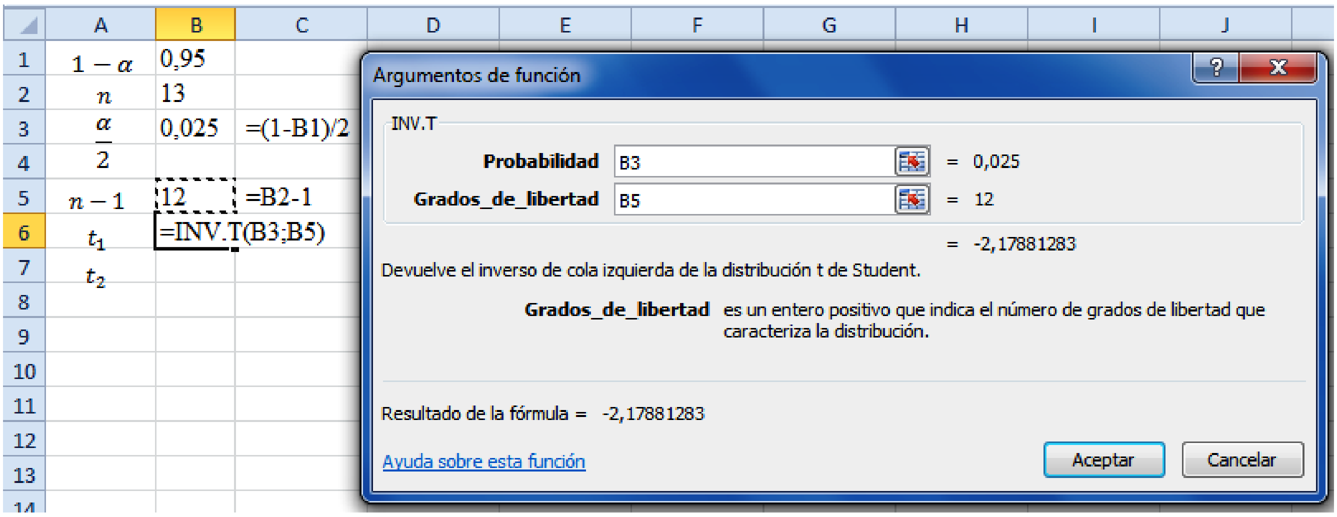
En la tabla con 12 grados de libertad y 0,025 de área se obtiene un valor de t =2,1788, y por simetría es igual también a t = -2,1788

Para realizar los cálculos en Excel se procede de la siguiente manera:

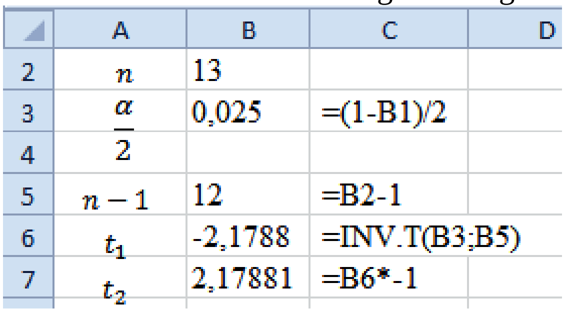
a) Llenar los datos y hacer los cálculos del área de una cola y de los grados de libertad. Luego insertar función. En la casilla seleccionar una categoría, seleccionar Estadísticas. Seleccionar la función INV.T.



b) Clic en Aceptar. En la ventana Argumentos de la función, en Probabilidad seleccionar B3, y en Grados de libertad seleccionar B6.



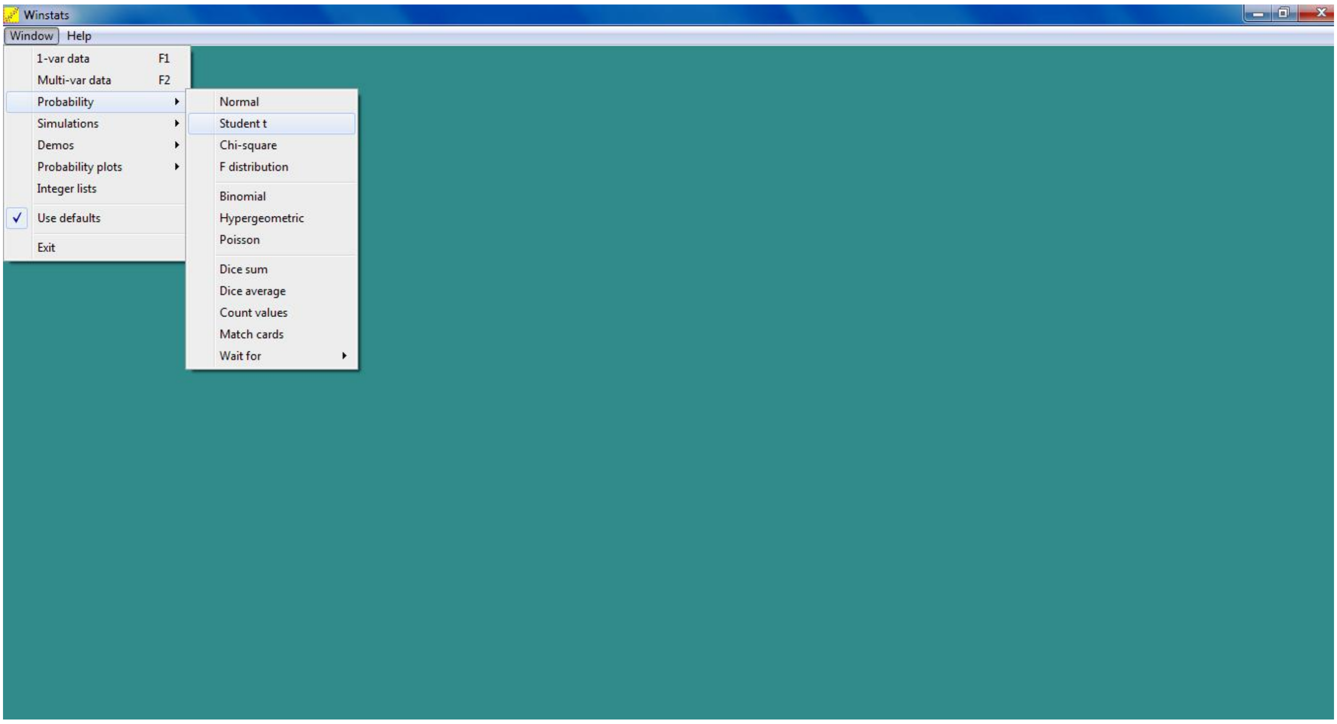
c) Clic en Aceptar. Los demás cálculos se muestran en la siguiente figura:



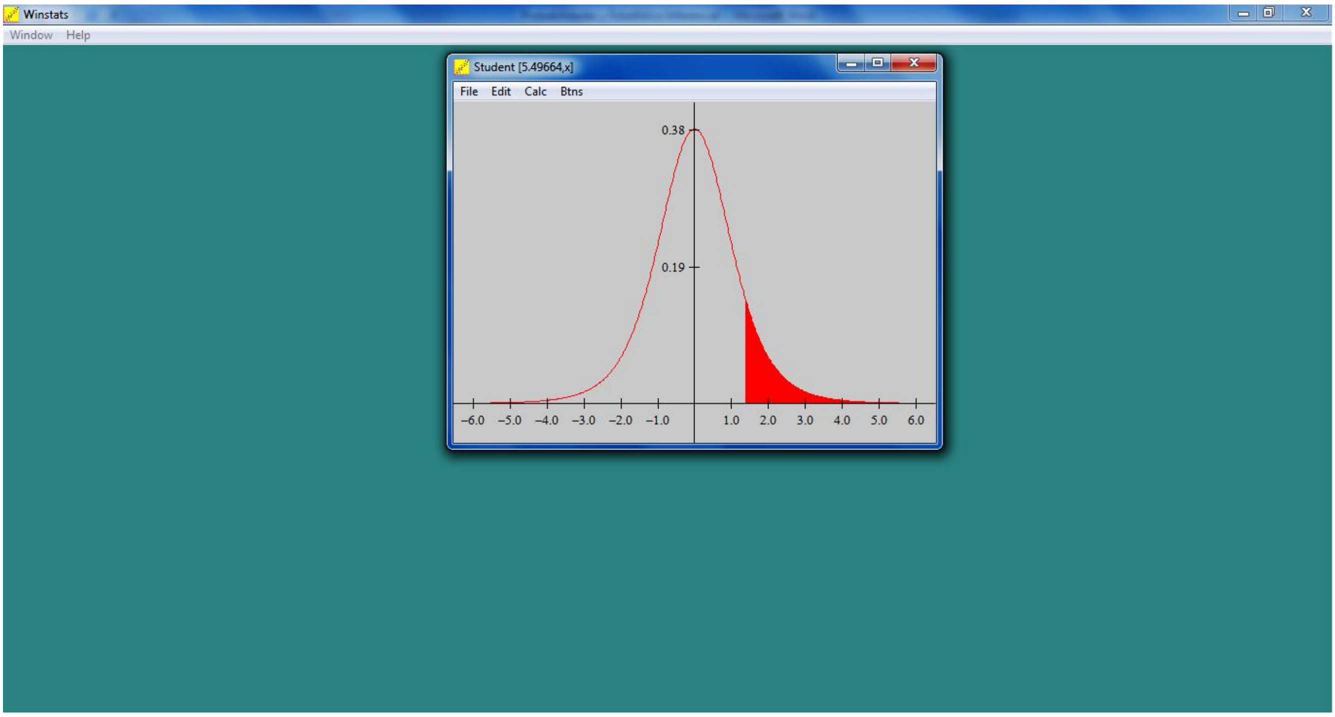
	A	B	C	D
2	n	13		
3	$\frac{\alpha}{2}$	0,025	=(1-B1)/2	
4				
5	$n-1$	12	=B2-1	
6	t_1	-2,1788	=INV.T(B3;B5)	
7	t_2	2,17881	=B6*-1	

Para resolver con Winstats se procede de la siguiente manera:

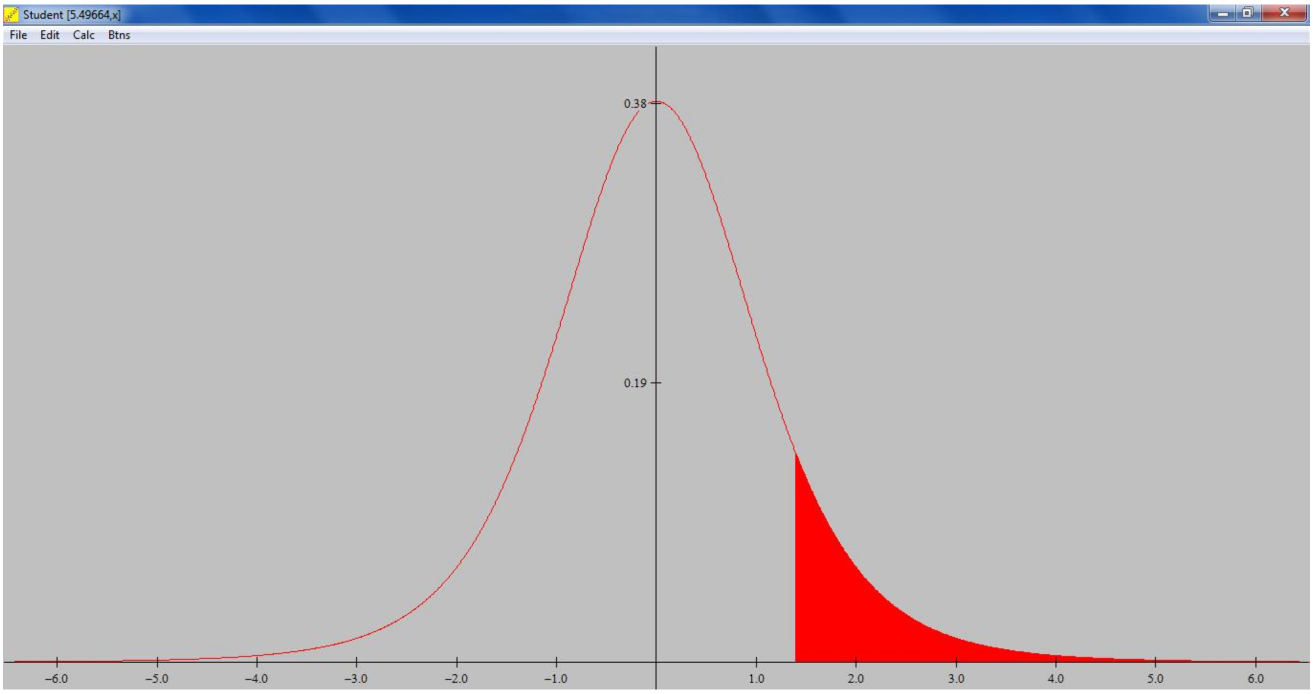
a) Clic en Window y luego en Probability seleccionar Student t



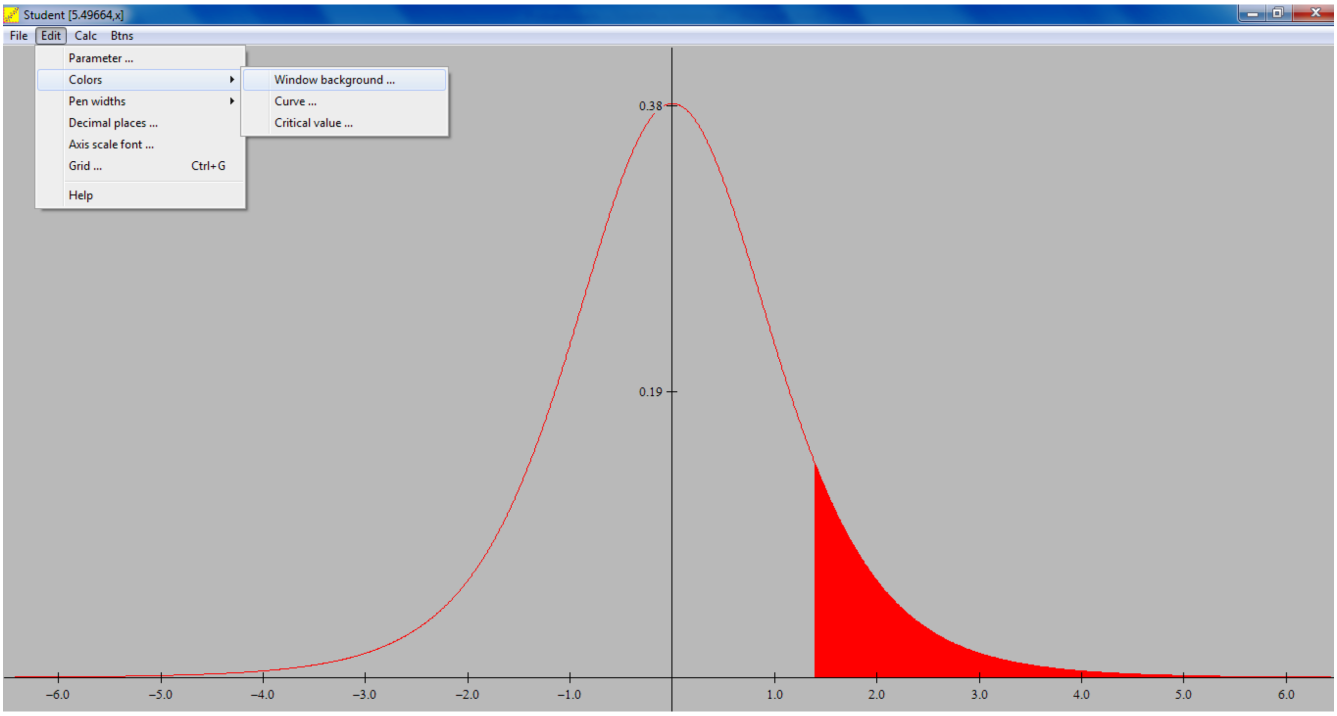
b) Clic en Student t



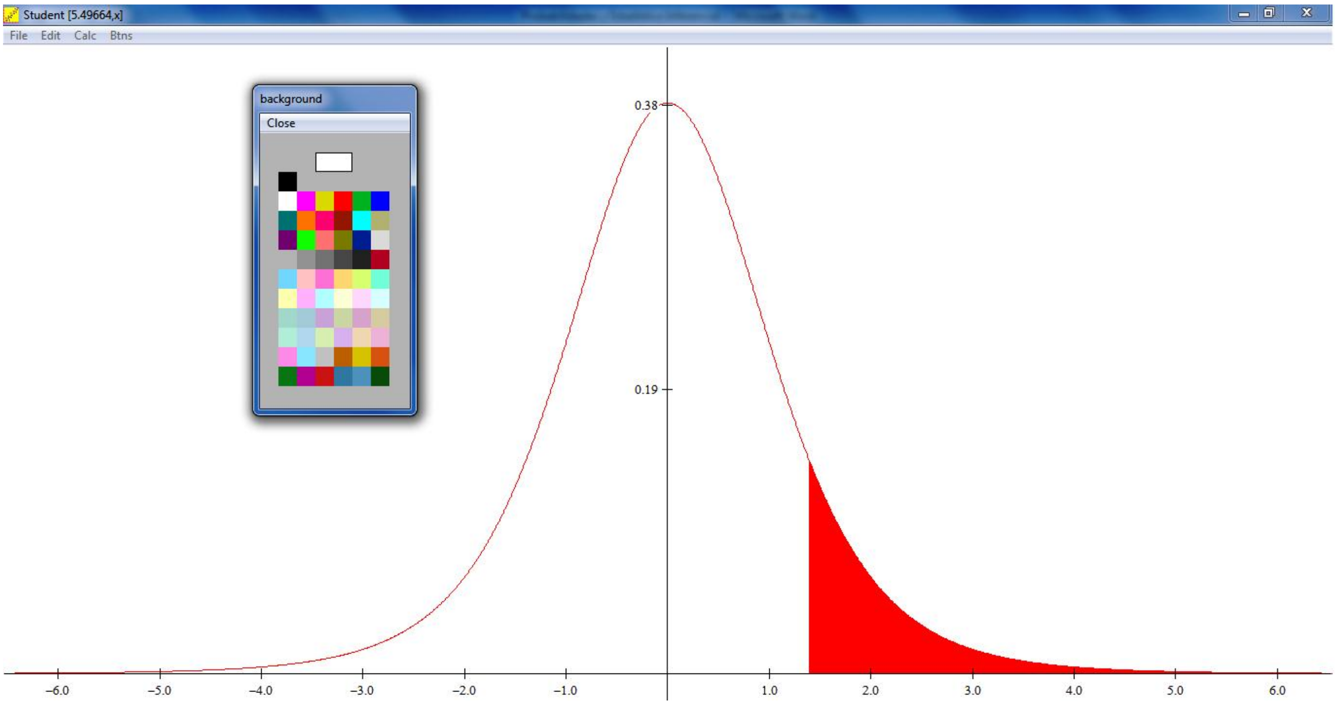
c) Maximizar la ventana de la distribución



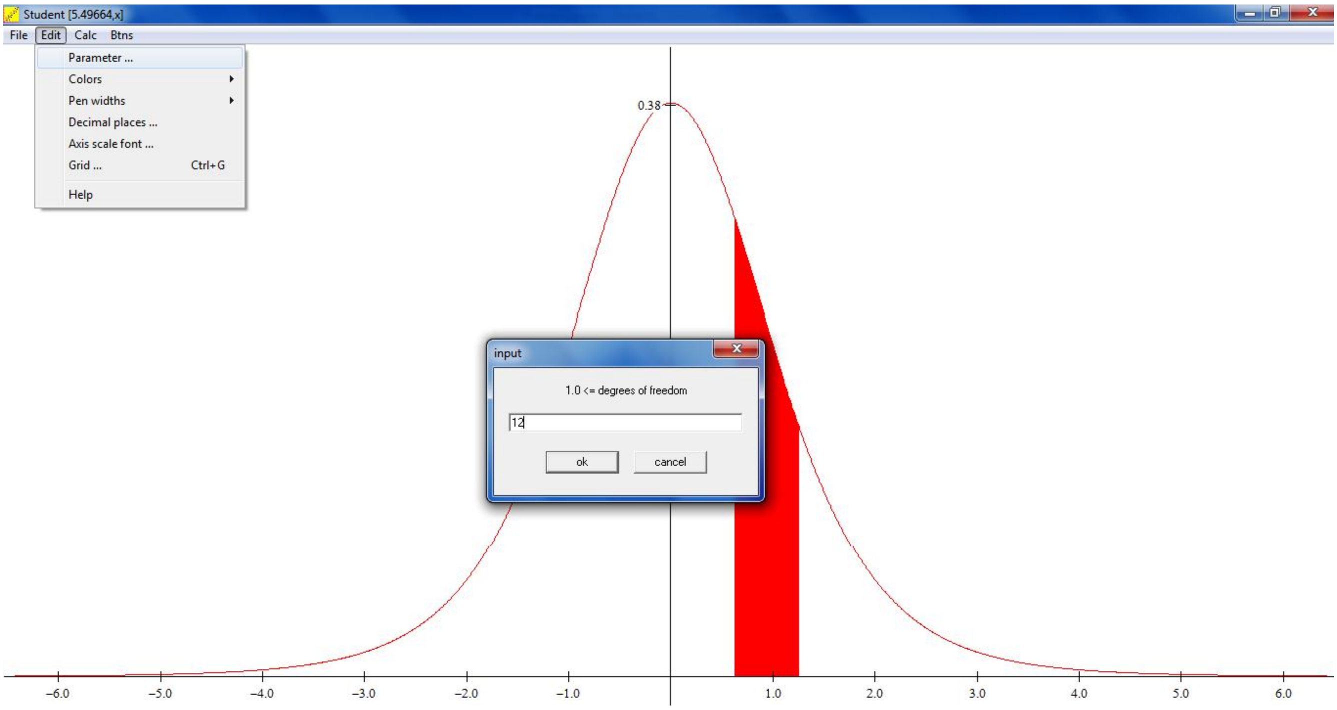
d) Para cambiar el color del fondo, clic en Edit + Colors + Window background



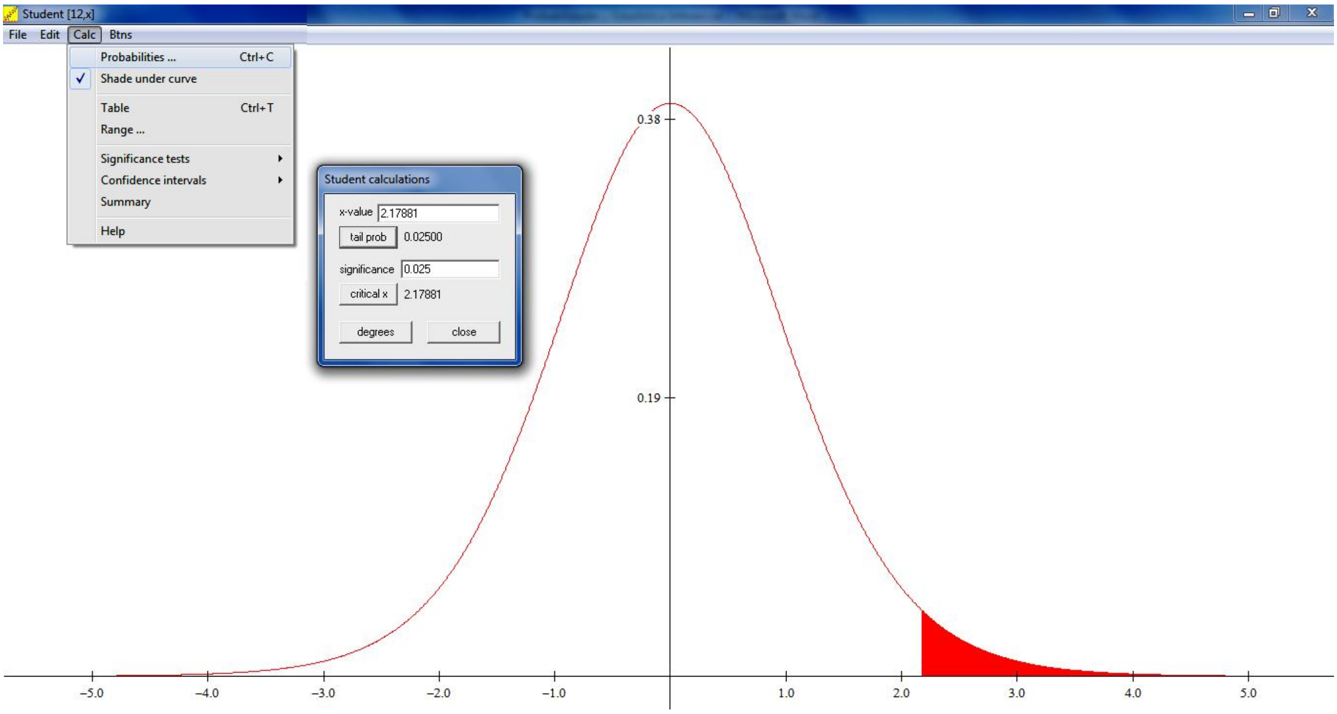
e) Clic Window background. En la venta de background seleccionar el color deseado, que este caso se seleccionó el color blanco. Luego clic en Close para cerrar la venta background.



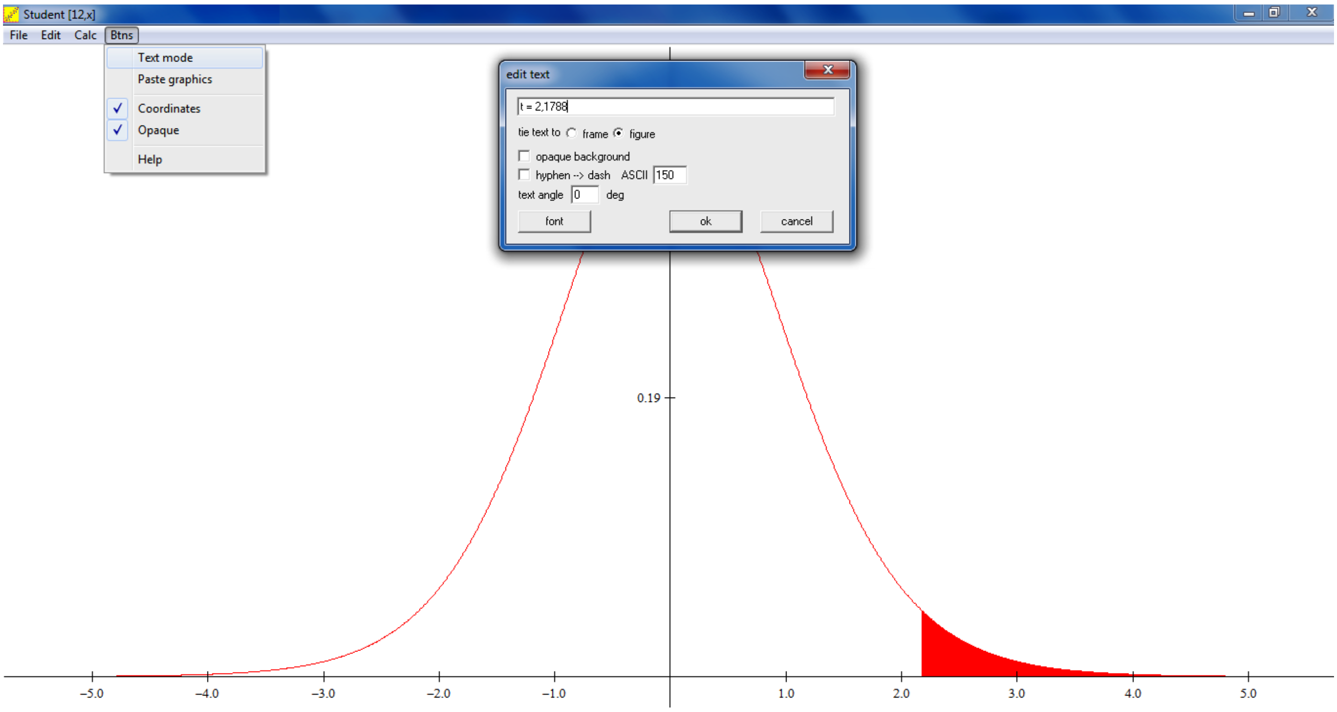
f) Para editar lo grados de libertad, clic en Edit + Parameter...(Parámetros). Clic en Parámetros. En la casilla de la ventana input escribir 12. Clic en ok



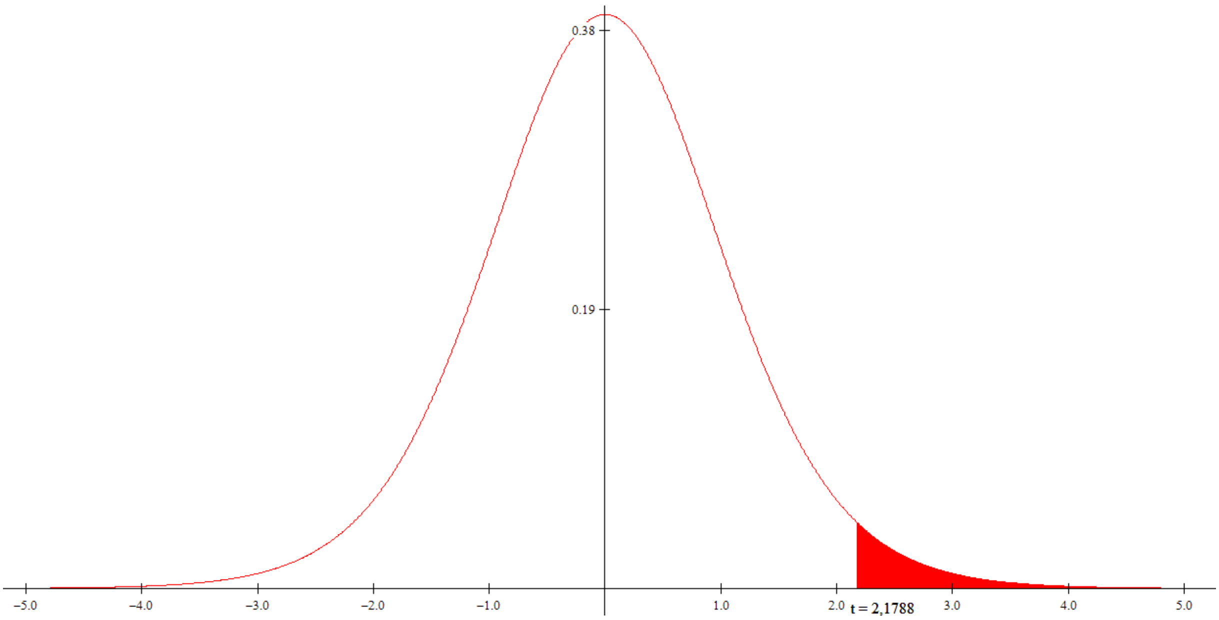
g) Para calcular el valor crítico de t, clic en Calc + Probabilities. En la ventana Student calculations, en significanse escribir 0,025 y luego clic en critical x. Clic en close para cerrar la ventana Student calculations.



h) Para escribir textos, clic en Btns. Luego clic derecho en cualquier parte de la ventana y aparece la ventana edit text. En la casilla de la ventana edit text escribir el texto deseado.



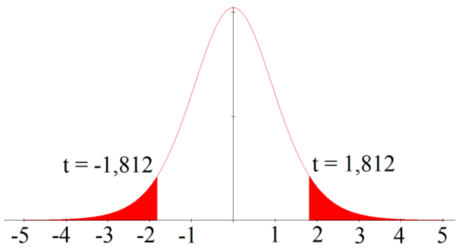
i) Clic en ok de la ventana edit text. Luego arrastar con el mause el texto al lugar deseado



2.2) Sea X = $t_{(10)}$ hallar el valor de P(X<-1,3722) + P(X> 2,7638) con lectura en la tabla, Excel y Winstats

Solución:

TABLA
DISTRIBUCIÓN t DE STUDENT



Ejemplos:
Para n-1 = 10 grados de libertad
$\boldsymbol{P(t > 1,812) = 0,05}$
$\boldsymbol{P(t < -1,812) = 0,05}$

n-1 \ α	0,25	0,2	0,15	**0,1**	0,05	0,025	**0,01**	0,005	0,0005
1	1,0000	1,3764	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567	636,6192
2	0,8165	1,0607	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248	31,5991
3	0,7649	0,9785	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409	12,9240
4	0,7407	0,9410	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041	8,6103
5	0,7267	0,9195	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321	6,8688
6	0,7176	0,9057	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074	5,9588
7	0,7111	0,8960	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995	5,4079
8	0,7064	0,8889	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554	5,0413
9	0,7027	0,8834	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498	4,7809
10	0,6998	0,8791	1,0931	**1,3722**	1,8125	2,2281	**2,7638**	3,1693	4,5869
11	0,6974	0,8755	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058	4,4370

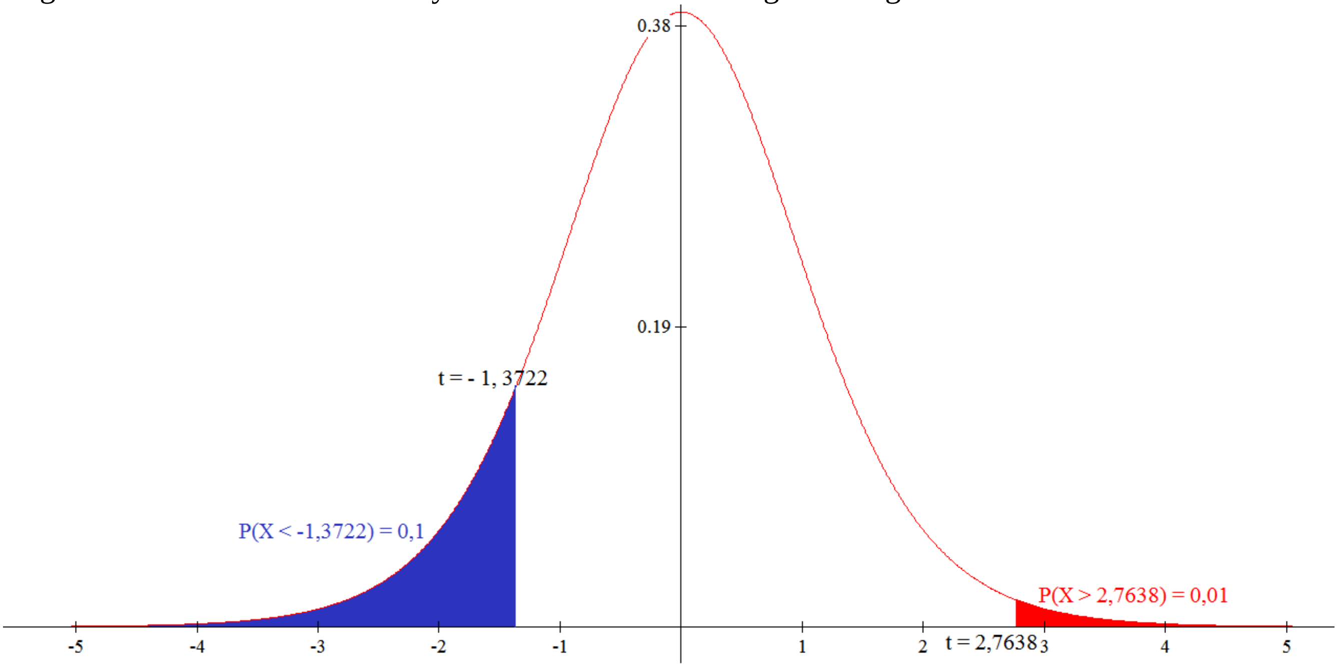
Con lectura en la tabla se obtiene:
P(X<-1,3722) = 0,1 y P(X > 2,7638) = 0,01

Entonces:
P(X<-1,3722) + P(X> 2,7638) = 0,1 + 0,01 = 0,11

Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E
1	t_1	-1,3722			
2	t_2	2,7638			
3	$n-1$	10			
4	$P(t_1 < -1{,}3722)$	0,1	=DISTR.T.N(B1;B3;VERDADERO)		
5	$P(t_2 > 2{,}7638)$	0,01	=DISTR.T.CD(B2;B3)		
6	$P(X < t_1) + P(X > t_2)$	0,11	=B4+B5		

El gráfico elaborado en Winstats y Paint se muestra en la siguiente figura:



2.3) Un fabricante de papel para computadora tiene un proceso de producción que opera continuamente a lo largo del turno. Se espera que el papel tenga una media de longitud de 11 pulgadas. De 500 hojas se selecciona una muestra de 29 hojas con una media de longitud del papel de 10,998 pulgadas y una desviación estándar de 0,02 pulgadas. Calcular la estimación del intervalo de confianza del 99%

Solución:

Los datos del problema son:

$\mu = 11$

$N = 500$

$n = 29$

$\bar{X} = 10{,}998$

$S = 0{,}02$

$Confianza = 99\%$

Como en los datos aparece el tamaño de la población, se debe verificar si el tamaño de la nuestra es mayor que el 5% para emplear la fórmula con el factor finito de corrección. Se remplaza valores en la siguiente fórmula:

$$\frac{n}{N} \cdot 100\% > 5\%$$

$$\frac{29}{500} \cdot 100\% = 5{,}8\%$$

Por lo tanto se debe utilizar la fórmula con el factor finito de corrección.

Calculando la proporción de la cola superior e inferior de la distribución se obtiene:

$$Nivel\ de\ confianza = (1 - \alpha) \cdot 100\%$$

$$\frac{\alpha}{2} = \frac{100\% - Nivel\ de\ confianza}{200}$$

$$\frac{\alpha}{2} = \frac{100\% - 99\%}{200} = 0{,}005$$

Calculando los grados de libertad se obtiene:

$n - 1 = 29 - 1 = 28$

Con lectura en la tabla para un área de 0,005 y 28 grados de libertad se obtiene t = ±2,7633

Remplazando valores y realizando lo cálculos se obtiene:

$$\bar{X} - t_{n-1}\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \leq \mu \leq \bar{X} + t_{n-1}\,\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$10{,}998 - 2{,}7633\frac{0{,}02}{\sqrt{29}}\sqrt{\frac{500-29}{500-1}} \leq \mu \leq 10{,}998 + 2{,}7633\frac{0{,}02}{\sqrt{29}}\sqrt{\frac{500-29}{500-1}}$$

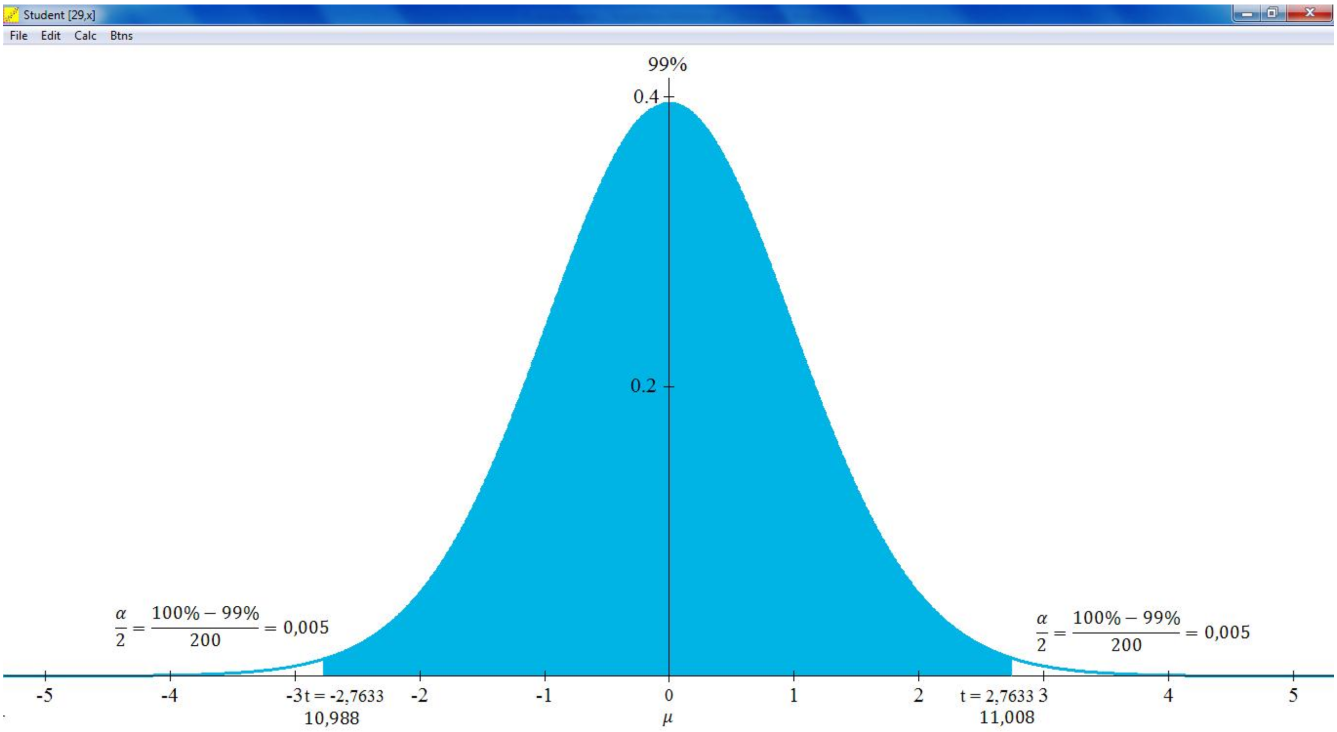
$$10{,}988 \leq \mu \leq 11{,}008$$

Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F
1	μ	11				
2	N	500				
3	n	29				
4	$\bar{X}$	10,998				
5	S	0,02				
6	Confianza	99				
7	$\frac{n}{N}\cdot 100 > 5\%$	5,8	=(B3/B2)*100			
8						
9	α	0,01	=(100-B6)/100			
10	$t_{n-1}\frac{S}{\sqrt{n}}$	0,0103	=INTERVALO.CONFIANZA.T(B9;B5;B3)			
11						
12						
13	$\bar{X} - t_{n-1}\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \leq \mu \leq \bar{X} + t_{n-1}\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$					
14						
15	10,988	$\leq \mu \leq$	11,008			
16	=B4-B10*RCUAD((B2-B3)/(B2-1))		=B4+B10*RCUAD((B2-B3)/(B2-1))			

Interpretación: Existe un 99% de confianza de que la media poblacional se encuentra entre 10,998 y 11,008

El gráfico elaborado en Winstats y Paint se muestra en la siguiente figura:



2.4) La duración media de lámparas producidas por una compañía han sido en el pasado de 1120 horas. Una muestra de 8 lámparas de la producción actual dio una duración media de 1070 horas con una desviación típica de 125 horas. Comprobar la hipótesis $\mu = 1120$ horas contra la hipótesis alternativa $\mu < 1200$ horas mediante un error tipo I de 0,05.

Solución:

Los datos son:
$\mu = 1120$
$n = 8$
$\bar{x} = 1070$
$S = 125$
$\alpha = 0{,}05$

Las hipótesis son:
$H_0: \mu = 1120$
$H_1: \mu < 1120$

Como se conoce la desviación estándar de la muestra S se debe utilizar la distribución t de Student. Con lectura en la tabla para un área de 0,025 y con $n - 1 = 8 - 7$ grados de libertad le corresponde un valor $t_{tabla} = -1{,}8946$. Se toma en cuenta el valor negativo porque se trata de una prueba de hipótesis a cola izquierda como se puede observar en la H_1.

Entonces para calcular el valor de t_{prueba} se emplea la siguiente ecuación:

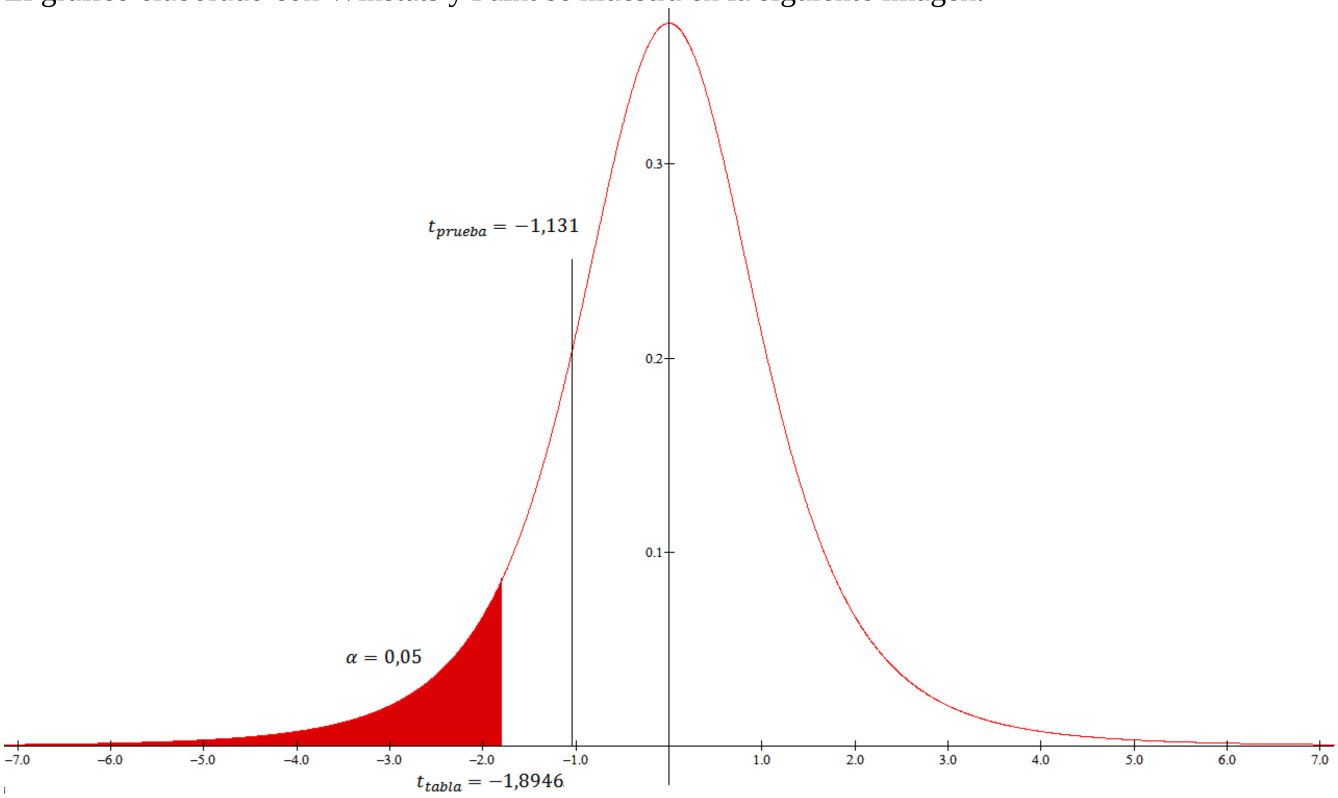
$$t_{prueba} = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

$$t_{prueba} = \frac{1070 - 1120}{\frac{125}{\sqrt{8}}} = -1{,}131$$

Los cálculos en Excel se muestran en la siguiente imagen:

	A	B	C	D
1	n	8		
2	$\bar{x}$	1070		
3	S	125		
4	μ	1120		
5	$H_0: \mu = 1120$			
6	$H_1: \mu < 1120$			
7	α	0,05		
8	n-1	7	=B1-1	
9	t_{tabla}	-1,8946	=INV.T(B7;B8)	
10	$t_{prueba} = \dfrac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$	-1,131	=(B2-B4)/(B3/RCUAD(B1))	
11				
12				

El gráfico elaborado con Winstats y Paint se muestra en la siguiente imagen:



Decisión: Dado que $-1{,}131 > t_{tabla} - 1{,}8946$ se Acepta la H_0

3) Tarea de interaprendizaje

1) ¿Cuál fue el origen del nombre de la distribución t de Student
2) Describa las diferencias entre la distribución t y la distribución normal
3) ¿En qué caso la distribución t es virtualmente idéntica a la distribución normal?
4) Defina con sus propias palabras el concepto de grados de libertad. De un ejemplo
5) Realice un organizador gráfico (cuadro sinóptico, mapa conceptual, etc.) sobre la distribución t de Student

6) Determinar el valor crítico de t con lectura en la tabla, Excel y Winstats en cada una de las siguientes condiciones:

6.1) $1-\alpha = 0{,}95$; n = 10

2,262

6.2) $1-\alpha = 0{,}99$; n = 10

3,250

7) Llenar la siguiente tabla de valores t para una confianza de 95% con lectura en la tabla, Excel y Winstats

Tamaño de muestra n	Grados de libertad n-1	Valor t
8		
13		
23		
28		

8) Sea X = $t_{(10)}$ hallar el valor de las siguientes probabilidades con lectura en la tabla, Excel y Winstats

8.1) P(X<-1,8125) + P(X> 1,8125)

0,1

8.2) P(X<-1,3722) + P(X> 1,3722)

0,2

8.3) P(X>2,2281)

0,025

8.4) P(X<- 2,2281)

0,025

8.5) P(X< 0,6998)

0,75

8.6) P(X< - 0,6998)

0,25

8.7) P (-1,0931<X<2,7638)

0,84

8.8) P (-0,6998<X<4,5869)

0,75

8.9) P (0,6998<X<4,5869)

0,2495

8.10) P (-0,6998<X<-4,5869)

0,2495

9) Si $\bar{X}$ = 50, S = 15, n=16, construya una estimación del intervalo de confianza del 99% de la media poblacional μ con lectura en la tabla, Excel y Winstats.

$38{,}95 \leq \mu \leq 61{,}05$

10) Si $\bar{X}$ = 40, S = 10, n=15, construya una estimación del intervalo de confianza del 90% de la media poblacional μ con lectura en la tabla, Excel y Winstats.

$35{,}45 \leq \mu \leq 44{,}55$

11) Si $\bar{X}$ = 60, S = 16, n=20 construya una estimación del intervalo de confianza del 80% de la media poblacional μ con lectura en la tabla, Excel y Winstats.

$55{,}25 \leq \mu \leq 64{,}75$

12) Si $\bar{X}$ = 30, S = 6, n=15 construya una estimación del intervalo de confianza del 99,9% de la media poblacional μ con lectura en la tabla, Excel y Winstats.

$23{,}59 \leq \mu \leq 36{,}41$

13) Determine el intervalo de confianza de 95% con lectura en la tabla, Excel y Winstats para los casos siguientes:

13.1) Si $\bar{X}$ = 15, S =2, n=16 y N= 200

$13{,}975 \leq \mu \leq 16{,}025$

13.2) Si $\bar{X}$ = 14, S =2, n=16 y N= 150

$12{,}975 \leq \mu \leq 15{,}025$

14) Una empresa manufacturera produce aislantes eléctricos. Si los aislantes se rompen al usarse, muy posiblemente tendremos un corto circuito. Para probar la fuerza de los aislantes, se lleva a cabo una prueba destructiva para determinar cuánta fuerza se requiere para romperlos. Se mide la fuerza observando cuántas libras se aplican al aislante antes de que se rompa. La siguiente tabla lista 30 valores de este experimento de la fuerza en libras requerida para romper el aislante. Construya una estimación del intervalo de confianza del 95% para la población media de fuerza requerida para romper al aislante empleando Excel y Winstats.

1870	1728	1656	1610	1634	1784	1522	1696	1592	1662
1866	1764	1734	1662	1734	1774	1550	1756	1762	1866
1820	1744	1788	1688	1810	1752	1680	1810	1652	1736

$1689{,}96 \leq \mu \leq 1756{,}84$

15) La duración media de las bombillas producidas por una compañía han sido en el pasado de 1000 horas. Una muestra de 28 bombillas de la producción actual dio una duración media de 1050 horas con una desviación típica de 120 horas. Comprobar la hipótesis $\mu = 1000$ horas contra la hipótesis alternativa $\mu < 1000$ horas mediante un nivel de significancia de 0,05.

Dado que $2{,}205 > t_{tabla} - 1{,}073$ se Acepta la H_0

16) Cuál sería el resultado del problema anterior si la muestra hubiese sido tomada de una población de 500 bombillas.

Dado que $2{,}14 > t_{tabla} - 1{,}073$ se Acepta la H_0

REFERENCIAS BIBLIOGRÁFICAS

SUÁREZ, Mario, (2012), Interaprendizaje de Probabilidades y Estadística Inferencial con Excel, Winstats y Graph, Primera Edición. Imprenta M & V, Ibarra, Ecuador.