DISTRIBUCIÓN NORMAL CON EXCEL Y WINSTATS

1) Reseña histórica

Abrahan De Moivre (1733) fue el primero en obtener la ecuación matemática de la curva normal. Kart Friedrich Gauss y Márquez De Laplece (principios del siglo diecinueve) desarrollaron más ampliamente los conceptos de la curva. La curva normal también es llamada curva de error, curva de campana, curva de Gauss, distribución gaussiana o curva de De Moivre.
Su altura máxima se encuentra en la media aritmética, es decir su ordenada máxima corresponde a una abscisa igual a la media aritmética. La asimetría de la curva normal es nula y por su grado de apuntamiento o curtosis se clasifica en mesocúrtica.

2) Ecuación

Su ecuación matemática de la función de densidad es:

$$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

Donde:
$\sigma = desviación\ estándar$
$\sigma^2 = varianza$
$\pi = 3{,}141592654 \ldots\ldots constante\ matemática$
$e = 2{,}7182818 \ldots\ldots constante\ matemática$
X = valor en el eje horizontal
Y = altura de la curva para cualquier valor de x
$\mu = media\ aritmética$

Cuando se expresa la variable x en unidades estándar (fórmula de estandarización)

$$Z = \frac{X-\mu}{\sigma}$$

La ecuación anterior es reemplazada por la llamada forma canónica, la cual es

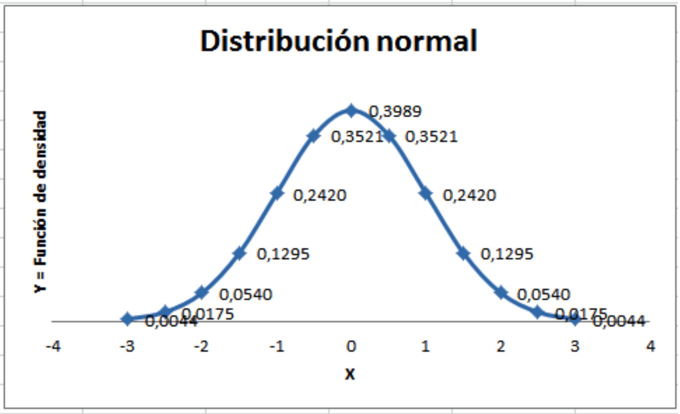
$$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}Z^2}$$

Para calcular Y en Excel se procede de la siguiente manera:

a) Se ubica valores para X del -3 hasta el 3. Se insertar la función DISTR.NORM.ESTAND.N. En la ventana de argumentos de función, en Z se seleccionada A2 que representa al -3, y en Acumulado es escribe FALSO. Clic en Aceptar. Se arrastra con el mouse para obtener los demás valores.

b) Para obtener la gráfica se inserta gráfico de dispersión.

	A	B	C
1	X	Y	$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$
2	-3	0,0044	=DISTR.NORM.ESTAND.N(A2;FALSO)
3	-2,5	0,0175	=DISTR.NORM.ESTAND.N(A3;FALSO)
4	-2	0,0540	=DISTR.NORM.ESTAND.N(A4;FALSO)
5	-1,5	0,1295	=DISTR.NORM.ESTAND.N(A5;FALSO)
6	-1	0,2420	=DISTR.NORM.ESTAND.N(A6;FALSO)
7	-0,5	0,3521	=DISTR.NORM.ESTAND.N(A7;FALSO)
8	0	0,3989	=DISTR.NORM.ESTAND.N(A8;FALSO)
9	0,5	0,3521	=DISTR.NORM.ESTAND.N(A9;FALSO)
10	1	0,2420	=DISTR.NORM.ESTAND.N(A10;FALSO)
11	1,5	0,1295	=DISTR.NORM.ESTAND.N(A11;FALSO)
12	2	0,0540	=DISTR.NORM.ESTAND.N(A12;FALSO)
13	2,5	0,0175	=DISTR.NORM.ESTAND.N(A13;FALSO)
14	3	0,0044	=DISTR.NORM.ESTAND.N(A14;FALSO)
15			



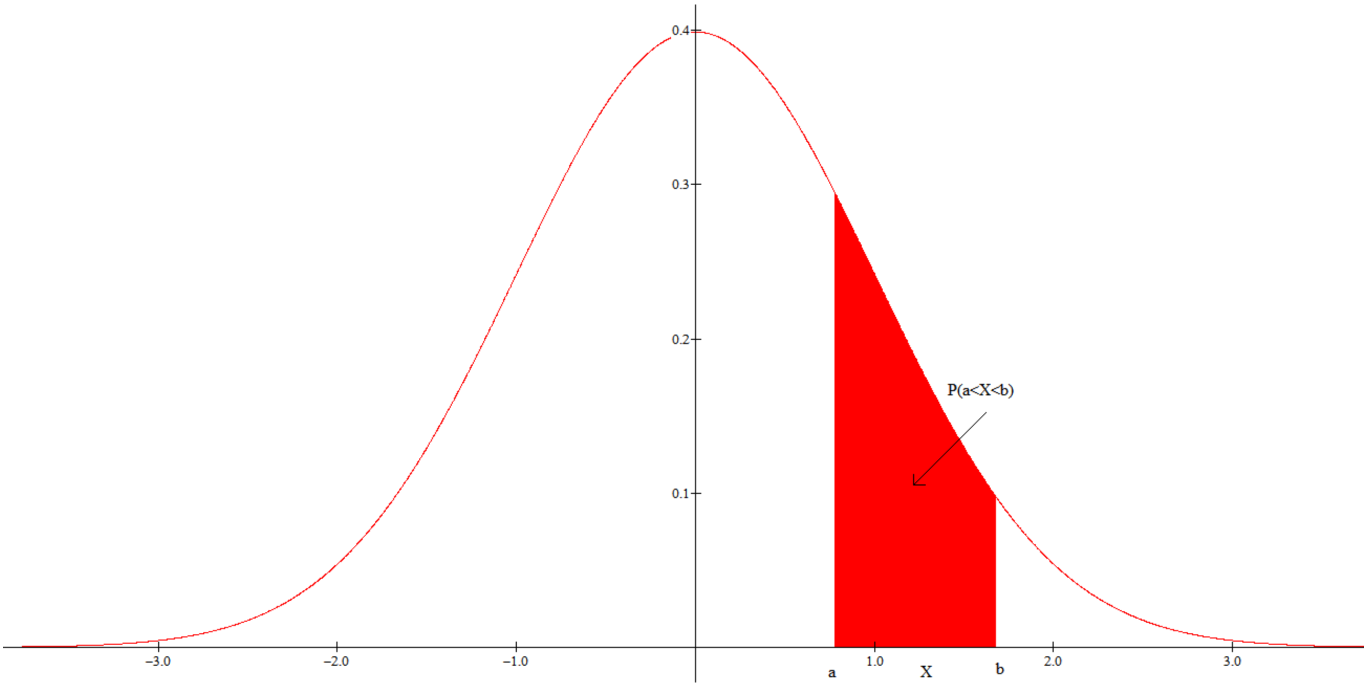
Nota: No existe una única distribución normal, sino una familia de distribuciones con una forma común, diferenciadas por los valores de su media y su varianza. De entre todas ellas, la más utilizada es la **distribución normal estándar**, que corresponde a una distribución con una media aritmética de 0 y una desviación típica de 1.

3) Área bajo la curva

El área total limitada por la curva y el eje "X" es 1, por lo tanto, el área bajo la curva entre X = a y X = b, con a < b, representa la probabilidad de que X esté entre a y b. Esta probabilidad se denota por:

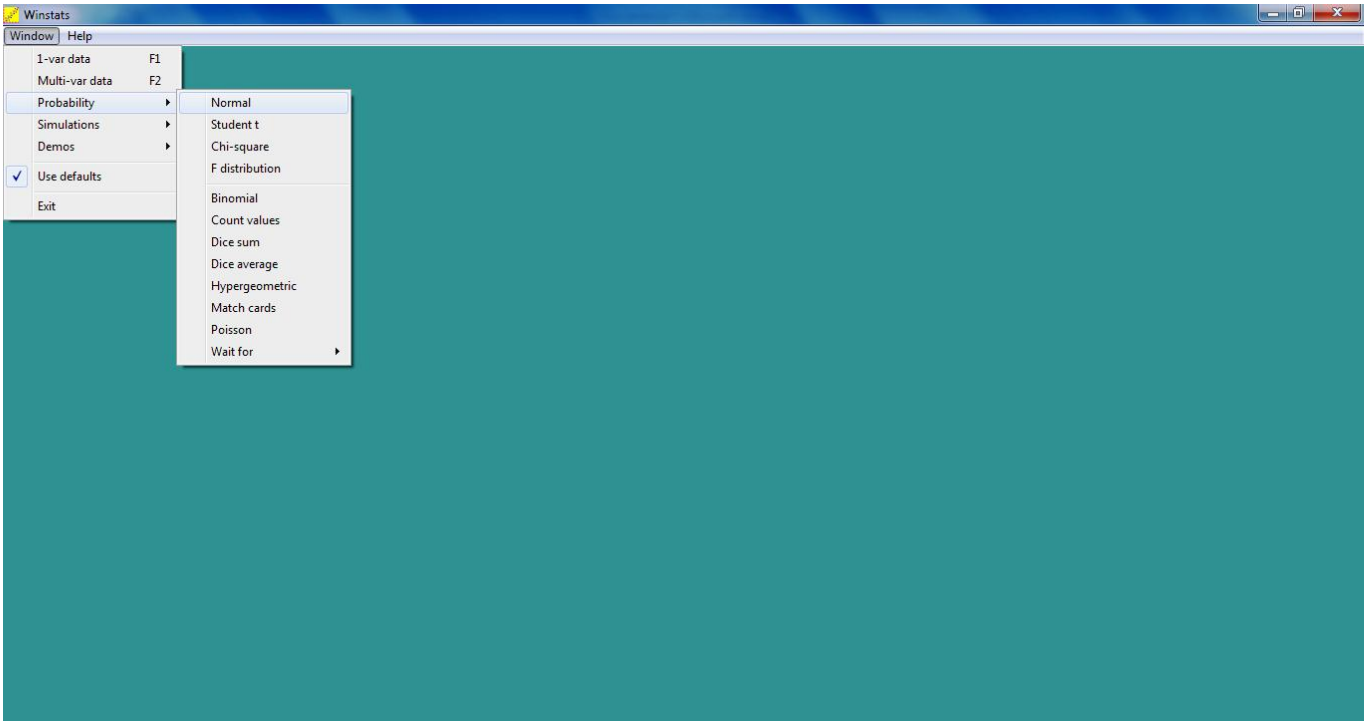
$$P(a < X < b)$$

Esta probabilidad se ilustra en el siguiente gráfico elaborado con el programa Winstats.

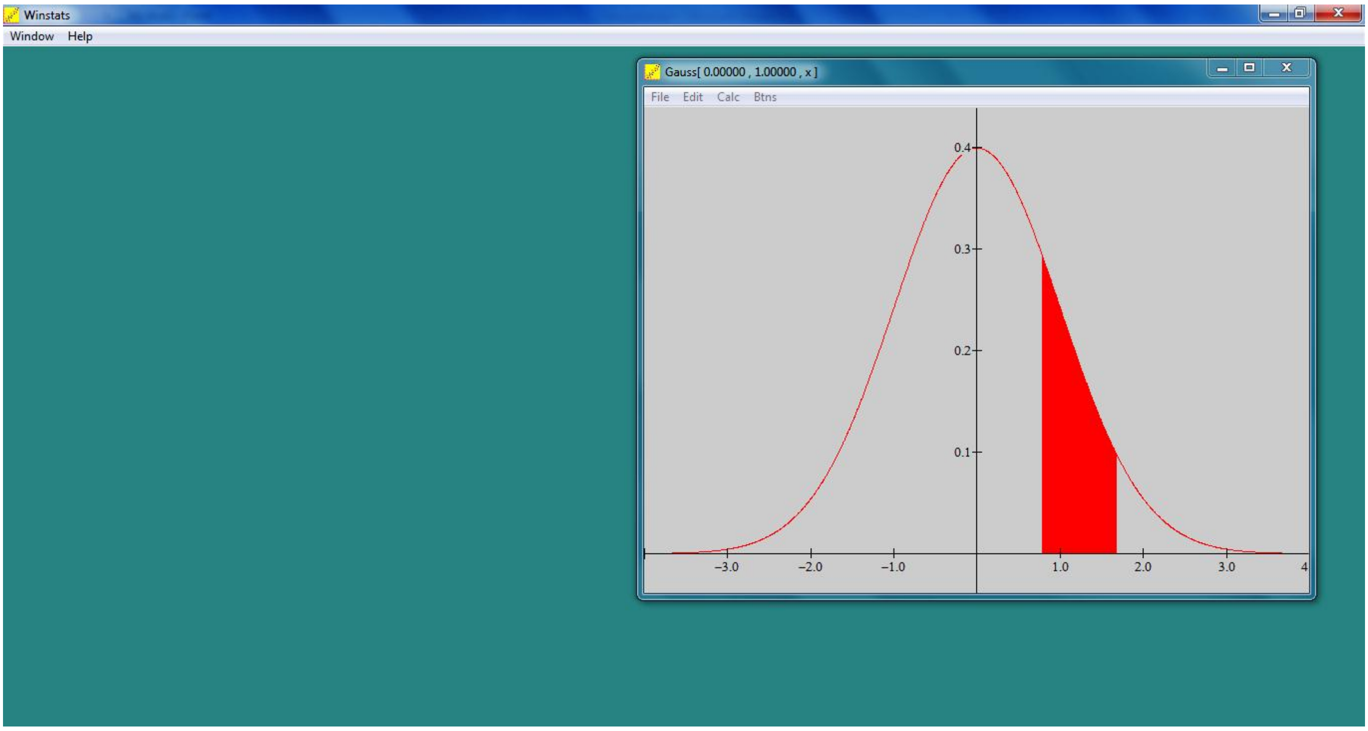


Para elaborar el gráfico se procede de la siguiente manera:

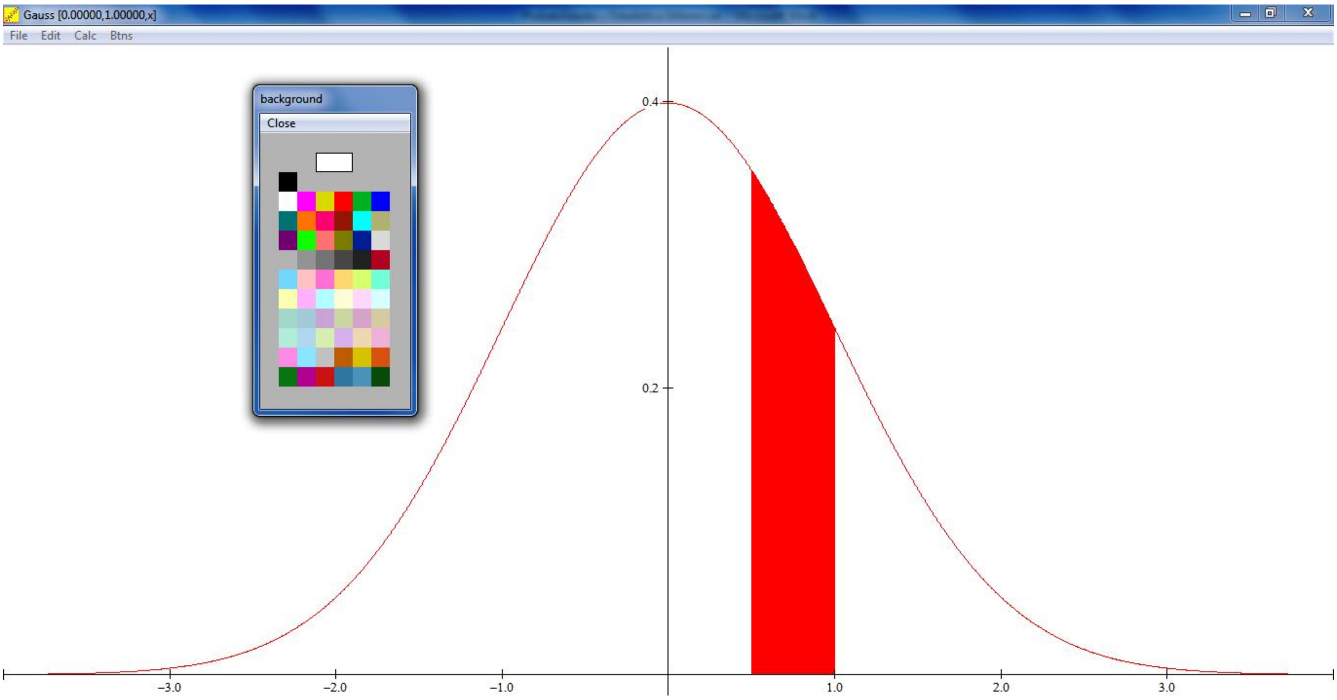
a) Se abre el programa. Clic en Window- Probability



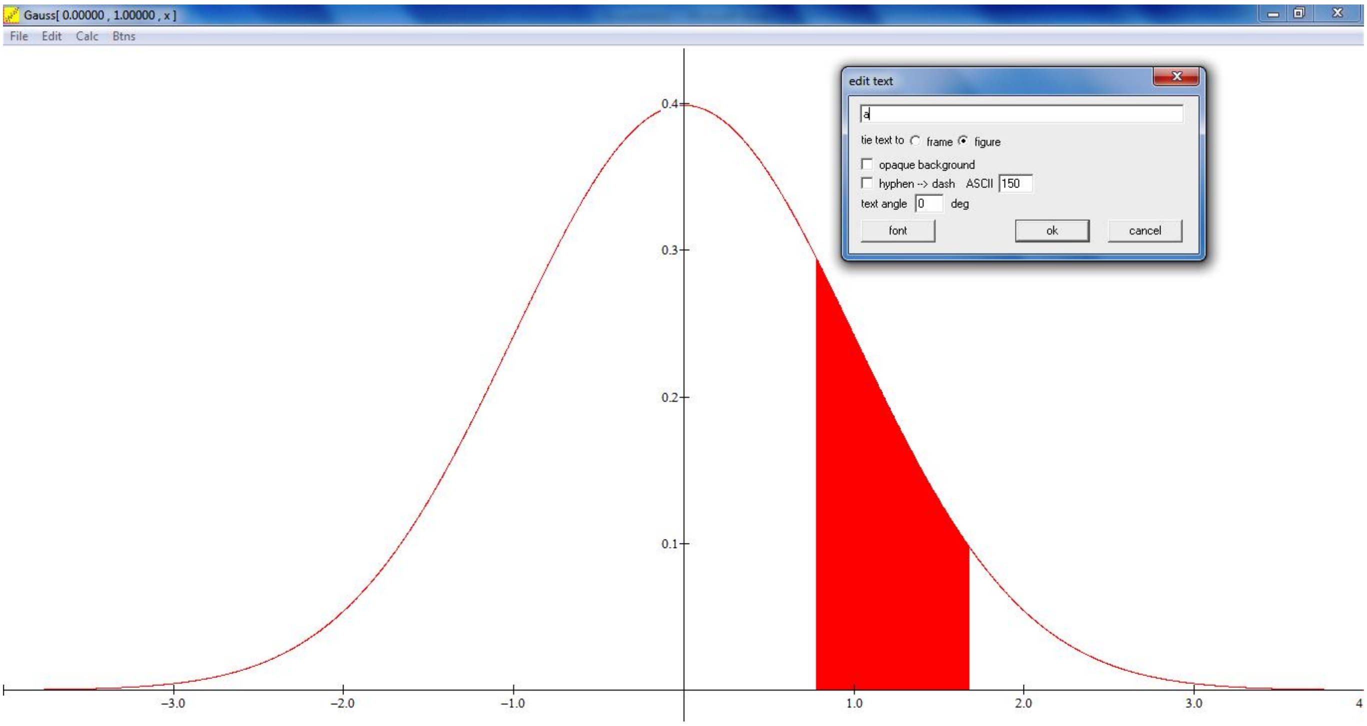
b) Clic en Normal



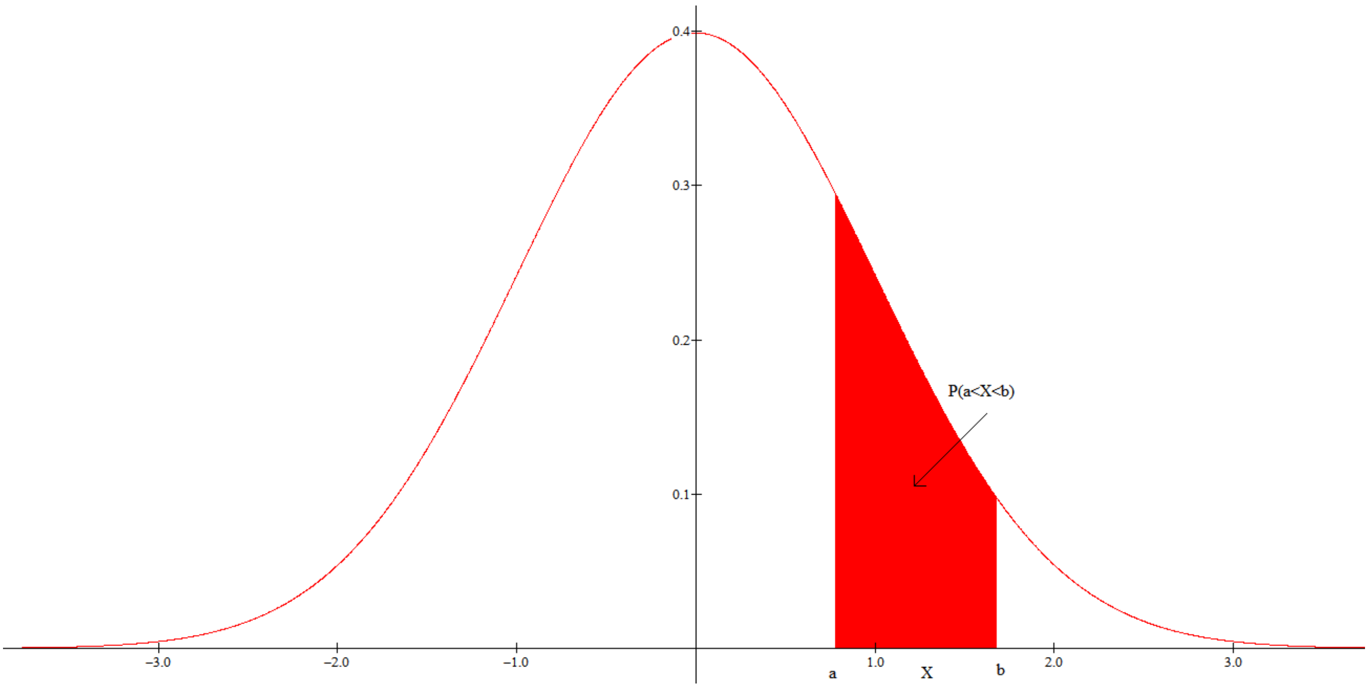
c) Para cambiar el color del fondo, maximizar la ventana de la curva. Clic en Edit-Colors y luego en Window background. Seleccionar el color blanco para el fondo.



d) Para escribir, clic en Btns y luego en Text mode. Clic derecho en cualquier parte de la pantalla. Luego escribir en la venta edit text. Clic en ok



e) Se obtiene el siguiente gráfico



4) Empleo de la distribución normal

La distribución normal suele emplearse en:

4.1) Estimación del intervalo de confianza para la media (σ conocida)

Se emplea la siguiente fórmula:

$$\bar{X} - Z\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z\frac{\sigma}{\sqrt{n}}$$

Donde:
Z = valor crítico de la distribución normal estandarizada

Se llama *valor crítico* al valor de Z necesario para construir un intervalo de confianza para la distribución. El 95% de confianza corresponde a un valor α de 0,05. El valor crítico Z correspondiente al área acumulativa de 0,975 es 1,96 porque hay 0,025 en la cola superior de la distribución y el área acumulativa menor a Z = 1,96 es 0,975.

Un nivel de confianza del 95% lleva a un valor Z de 1,96. El 99% de confianza corresponde a un valor α de 0,01. El valor de Z es aproximadamente 2,58 porque el área de la cola alta es 0,005 y el área acumulativa menor a Z = 2,58 es 0,995.

4.2) Estimación del intervalo de confianza para una proporción

Sirve para calcular la estimación de la proporción de elementos en una población que tiene ciertas características de interés. La proporción desconocida de la población, se representa con la letra griega π. La estimación puntual para π es la proporción de la muestra, $p = \frac{X}{n}$, donde n es el tamaño de la muestra y X es el número de elementos en la muestra que tienen la característica de interés. La siguiente ecuación define la estimación del intervalo de confianza para la proporción de la población.

$$p - Z\sqrt{\frac{p(1-p)}{n}} \leq \pi \leq p + Z\sqrt{\frac{p(1-p)}{n}}$$

Donde:

$$p = proporción\ de\ la\ muestra = \frac{X}{n} = \frac{número\ de\ elementos\ con\ característica\ de\ interés}{tamaño\ de\ la\ muestra}$$
$$\pi = proporción\ de\ la\ población$$
$$Z = valor\ crítico\ para\ la\ distribución\ normal\ estandarizada$$
$$n = tamaño\ de\ la\ muestra$$

Cuando la población es finita (N) y el tamaño de la muestra (n) constituye más del 5% de la población, se debe usar el factor finito de corrección. Por lo tanto si cumple:

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

Se aplica la ecuación

$$p - Z\sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}} \leq \pi \leq p + Z\sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$$

4.3) Cálculo del tamaño de la muestra
Se aplica la fórmula

$$n = \frac{N\sigma^2 Z^2}{(N-1)e^2 + \sigma^2 Z^2}$$

4.4) Prueba medias de una muestra

Se utiliza una prueba de una muestra para probar una afirmación con respecto a una media de una población única.

Si se conoce la desviación estándar de la población (σ), la distribución de muestreo adecuada es la distribución normal. Si la población que se muestra es normal, la distribución de muestreo será normal en el caso de todos los tamaños de la muestra, y el valor estadístico de prueba a utilizar es:

$$Z_{prueba} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

La anterior ecuación se aplican para poblaciones infinitas, pero cuando la población es finita y el tamaño de la muestra n constituye más del 5% del tamaño de la población N, es decir:

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

En este caso se debe usar el factor finito de corrección para modificar las desviaciones estándar, por lo tanto se aplican la siguiente ecuación

$$Z_{prueba} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}}$$

4.5) Prueba medias de dos muestras

Las pruebas de dos muestras se utilizan para decidir si las medias de dos poblaciones son iguales. Se requieren dos muestras independientes, una de cada una de las dos poblaciones. Considérese, por ejemplo, una compañía investigadora que experimentan con dos diferentes mezclas de pintura, para ver si se puede modificar el tiempo de secado de una pintura para uso doméstico. Cada mezcla es probada un determinado número de veces, y comparados posteriormente los tiempos medios de secado de las dos muestras. Una parece ser superior, ya que su tiempo medio de secado (muestra) es 30 minutos menor que el de la otra muestra.

Pero, ¿son realmente diferentes los tiempos medios de secado de las dos pinturas, o esta diferencia muestral es nada más la variación aleatoria que se espera, aun cuando las dos fórmulas presentan idénticos tiempos medios de secado? Una vez más, las diferencias casuales se deben distinguir de las diferencias reales.

Con frecuencia se utilizan pruebas de dos muestras para comparar dos métodos de enseñanza, dos marcas, dos ciudades, dos distritos escolares y otras cosas semejantes.

La hipótesis nula puede establecer que las dos poblaciones tienen medias iguales:

$$H_0: \mu_1 = \mu_2$$

Las alternativas pueden ser alguna de las siguientes:

$$H_1: \mu_1 \neq \mu_2 \qquad H_1: \mu_1 > \mu_2 \qquad H_1: \mu_1 < \mu_2$$

Cuando se conocen las desviaciones estándar de la población σ_1 y σ_1, el valor estadístico de prueba es el siguiente:

$$Z_{prueba} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

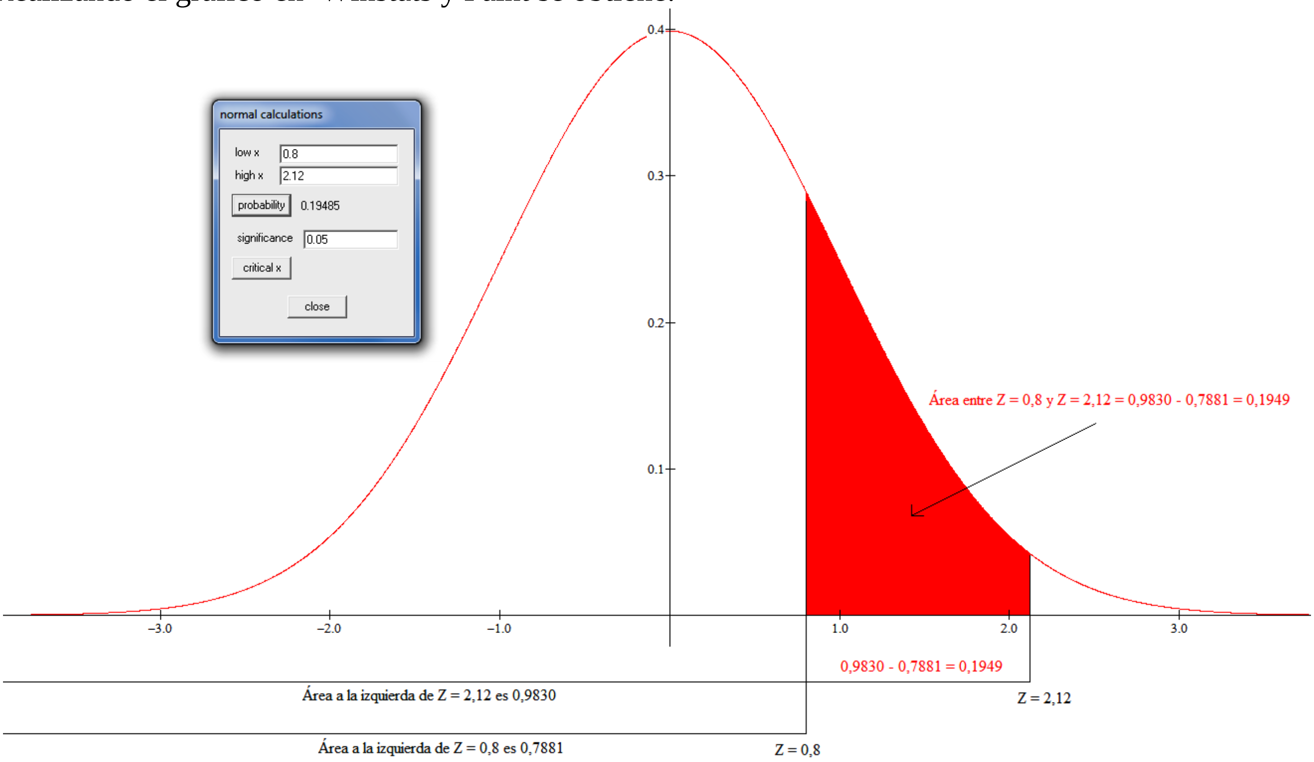
Cabe suponer que el valor real de Z, cuando H_0 es verdadera, está distribuido normalmente con una media de 0 y una desviación estándar de 1 (es decir, la distribución normal estandarizada) para casos en los que la suma $n_1 + n_2$ es igual o mayor de 30. Para tamaños más pequeños de muestra, Z estará distribuida normalmente sólo si las dos poblaciones que se muestrean también lo están.

5) Ejemplos ilustrativos

1) Averigüe el área bajo la curva de distribución normal entre Z = 0,8 y Z = 2,12

Solución:

Realizando el gráfico en Winstats y Paint se obtiene:



El área a la izquierda de Z = 0,8 con lectura en la tabla de la distribución normal es 0,7881
El área a la izquierda de Z = 2,12 con lectura en la tabla de la distribución normal es 0,9830

TABLA
DISTRIBUCIÓN NORMAL

$$Z = \frac{X-\mu}{\sigma}$$

Ejemplo:
$P(Z < -1{,}96) = 0{,}0250$

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
	⋮		⋮							
0,8	**0,7881**	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	**0,9830**	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857

El área Z = 0,8 y Z = 2,12 es 0,9830 – 0,7881 = 0,1949

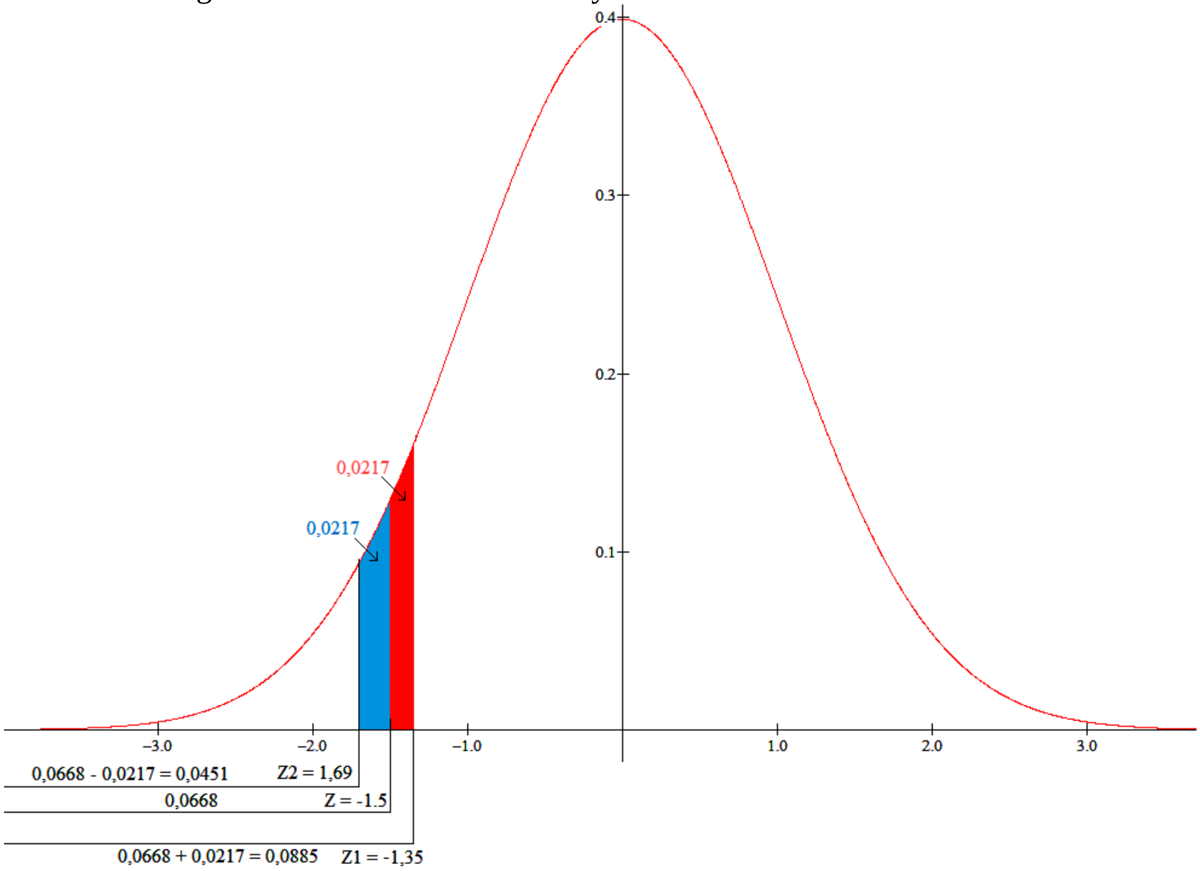
Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F
1	Z_1	0,8				
2	Z_2	2,12				
3	$\acute{A}rea\ a\ la\ izquierda\ de\ Z_1$	0,7881	=DISTR.NORM.ESTAND.N(B1;VERDADERO)			
4	$\acute{A}rea\ a\ la\ izquierda\ de\ Z_2$	0,9830	=DISTR.NORM.ESTAND.N(B2;VERDADERO)			
5	$P(0{,}8 < X < 2{,}12)$	0,1949	=B4-B3			

2) Halle Z si el área entre -1,5 y Z es 0,0217

Solución:

Realizando un gráfico ilustrativo en Winstats y Paint se obtiene:



Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E
1	Z	-1,5			
2	$Área\ entre\ Z\ y\ Z_1$	0,0217			
3	$Área\ a\ la\ izquierda\ de\ Z$	0,0668	=DISTR.NORM.ESTAND.N(B1;VERDADERO)		
4	$Área\ a\ la\ izquierda\ de\ Z_1$	0,0885	=B2+B3		
5	Z_1	-1,35	=INV.NORM.ESTAND(B4)		
6	$Área\ a\ la\ izquierda\ de\ Z_2$	0,0451	=B3-B2		
7	Z_2	-1,69	=INV.NORM.ESTAND(B6)		

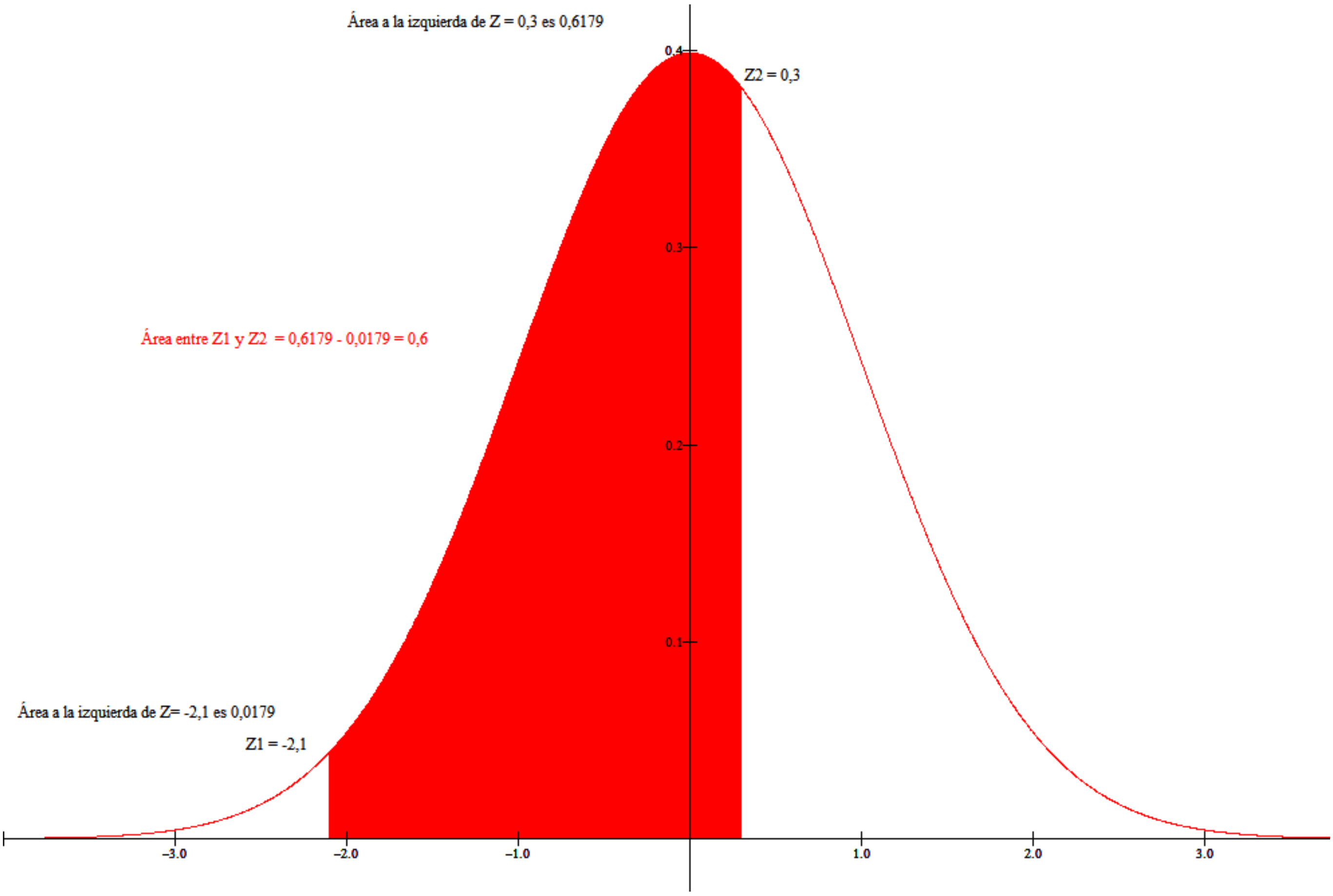
3) El peso de 200 estudiantes varones de cierta universidad es 151 libras, y la desviación típica es 15 libras. Si los pesos están distribuidos normalmente, calcular la probabilidad y el número de estudiantes que pesan Entre 120 y 155 libras

Solución: La curva normal corresponde a una función continua (valor decimal). Para resolver estos problemas se emplea los límites inferior y superior según sea el caso, es decir, para este problema es entre 119,5 y 155,5 libras

Normalizando los datos se tiene:

$$Z = \frac{X - \mu}{\sigma}$$
$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{119{,}5 - 151}{15} = -2{,}1$$
$$Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{155{,},5 - 151}{15} = 0{,}3$$

Graficando se obtiene:



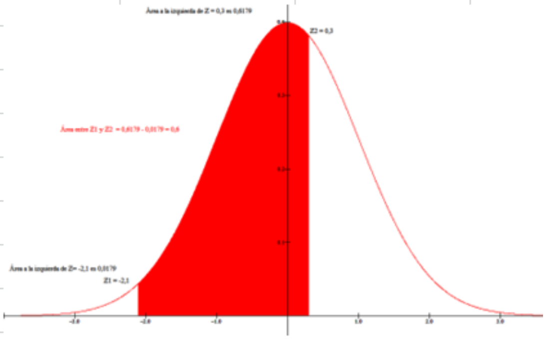
El área a la izquierda de Z = 0,3 con lectura en la tabla de la distribución normal es 0,6179
El área a la izquierda de Z = -2,1 con lectura en la tabla de la distribución normal es 0,0179

El área entre -2,1 y 0,3 es 0,6179 – 0,0179 = 0,6 = 60%

El número de estudiantes es 0,6 x 200 = 120

Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F	G
1	n	200					
2	μ	151					
3	σ	15					
4							
5	X_1	119,5	$Z = \frac{X - \mu}{\sigma}$				
6	X_2	155,5					
7	Z_1	-2,1	=NORMALIZACION(B5;B2;B3)				
8	Z_2	0,3	=NORMALIZACION(B6;B2;B3)				
9	$\acute{A}rea\ a\ la\ izquierda\ de\ Z_1$			0,0179	=DISTR.NORM.ESTAND.N(B7;VERDADERO)		
10	$\acute{A}rea\ a\ la\ izquierda\ de\ Z_2$			0,6179	=DISTR.NORM.ESTAND.N(B8;VERDADERO)		
11	$\acute{A}rea\ entre\ Z_1 y Z_2$			60%	=D10-D9		
12	Número de estudiantes			120	=D11*B1		

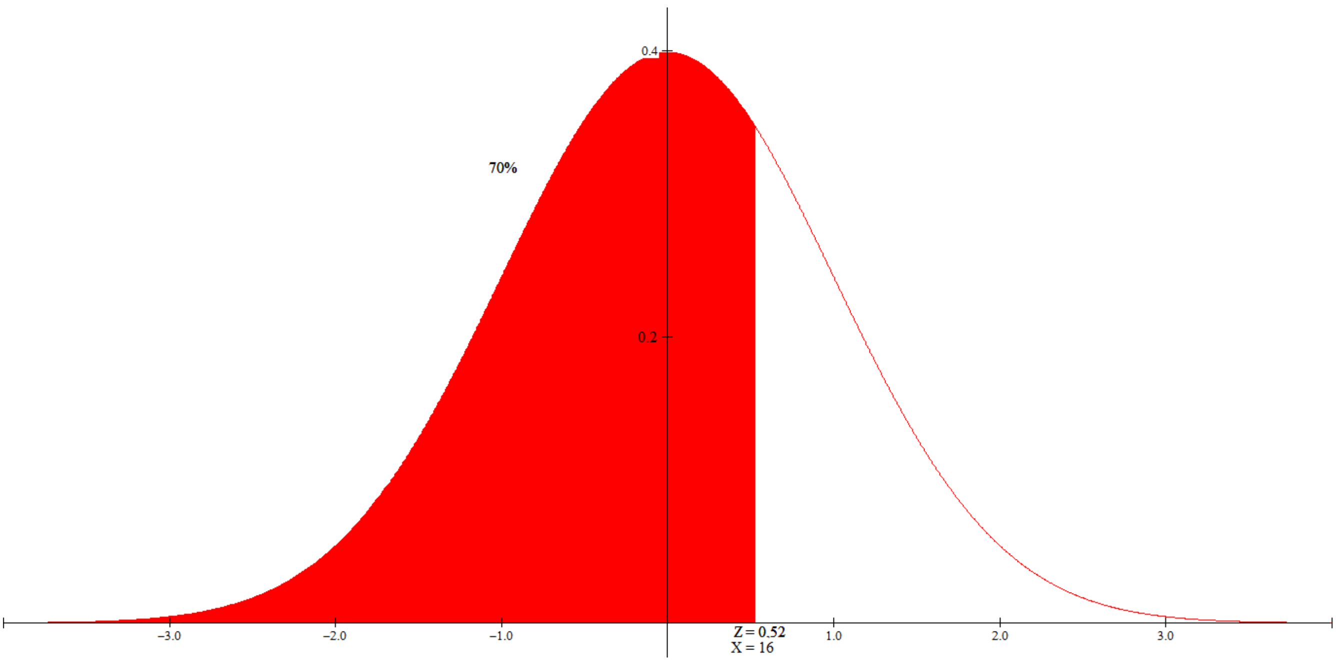


4) Las calificaciones que obtienen los alumnos en un examen siguen una distribución normal, siendo la media igual a 14. El 70% de los alumnos obtienen una calificación inferior a 16.
4.1) Calcule la desviación típica de las calificaciones
4.2) Se escoge un alumno al azar, calcule el porcentaje de obtener una calificación superior a 18

Solución
4.1)
Si el área es inferior al 70%, entonces con lectura en la tabla se obtiene el valor de Z = 0,52
Reemplazando valores en la fórmula y realizando las operaciones se obtiene:

$$Z = \frac{X - \mu}{\sigma} \Rightarrow 0{,}52 = \frac{16 - 14}{\sigma} \Rightarrow \sigma = \frac{16 - 14}{0{,}52} = 3{,}8$$

El gráfico elaborado en Winstats es:



4.2)

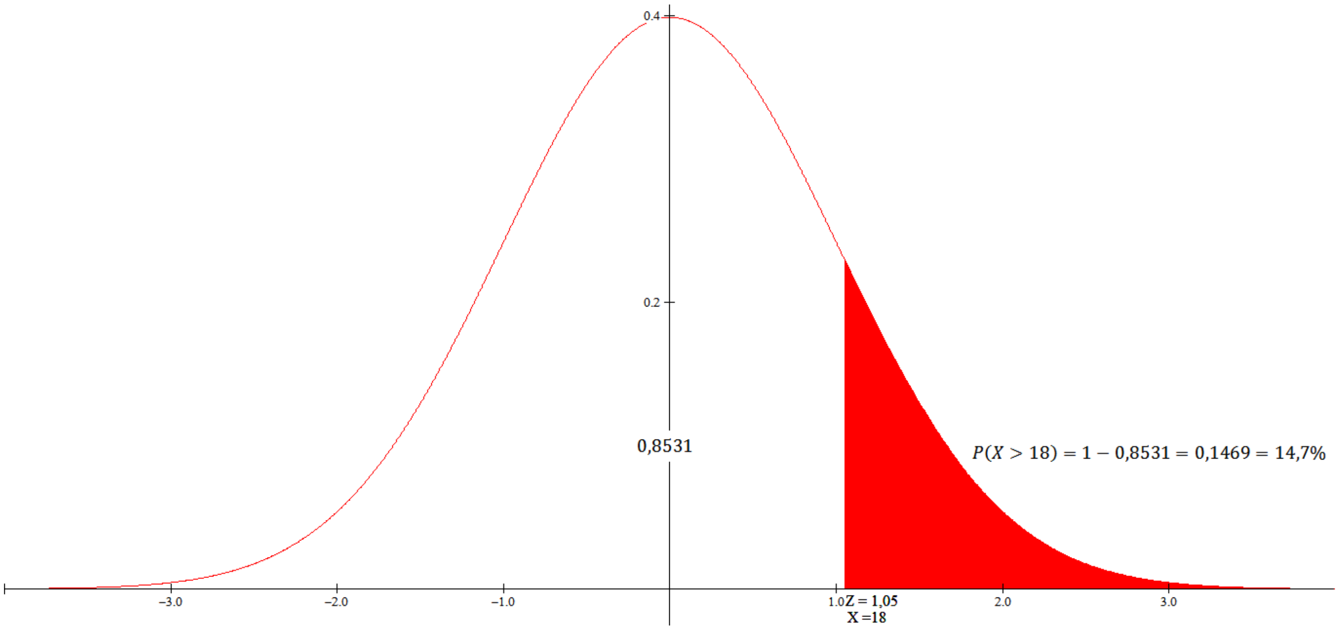
Reemplazando valores en la fórmula se obtiene el siguiente número Z:

$$Z = \frac{X - \mu}{\sigma} \Rightarrow Z = \frac{18 - 14}{3{,}8} \Rightarrow Z = 1{,}05$$

Con lectura en la tabla para Z = 1,05 se obtiene un área de 0,8531, la cual representa una probabilidad inferior a la calificación de 18

Para calcular la probabilidad de obtener una calificación superior a 18 se realiza la siguiente operación:
$P(X > 18) = 1 - 0{,}8531 = 0{,}1469 = 14{,}7\%$

El gráfico elaborado en Winstats y en Paint es:



Por lo tanto existe una probabilidad de 14,7% de obtener una calificación superior a 18

Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F
1	μ	14				
2	X	16				
3	Área a la izquierda de Z	0,7				
4	Z	0,52	=INV.NORM.ESTAND(B3)			
5	$Z = \frac{X - \mu}{\sigma} \Rightarrow \sigma = \frac{X - \mu}{Z}$	3,8	=(B2-B1)/B4			
6						
7						
8	X	18				
9	$Z = \frac{X - \mu}{\sigma}$	1,05	=(B8-B1)/B5			
10						
11	Área a la izquierda de Z	0,853	=DISTR.NORM.ESTAND.N(B9;VERDADERO)			
12	Porcentaje superior a 18	14,7%	=1-B11			

5) Calcular la proporción de cola superior e inferior para un intervalo del 95% de confianza

Solución:

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

Remplazando valores en la fórmula anterior del novel de confianza se obtiene:

$95\% = (1-\alpha)\cdot 100\% \Rightarrow \frac{95\%}{100\%} = (1-\alpha) \Rightarrow \alpha = 1 - \frac{95\%}{100\%} \Rightarrow \alpha = 1 - 0{,}95 = 0{,}05$

La proporción de la cola superior e inferior de la distribución es:

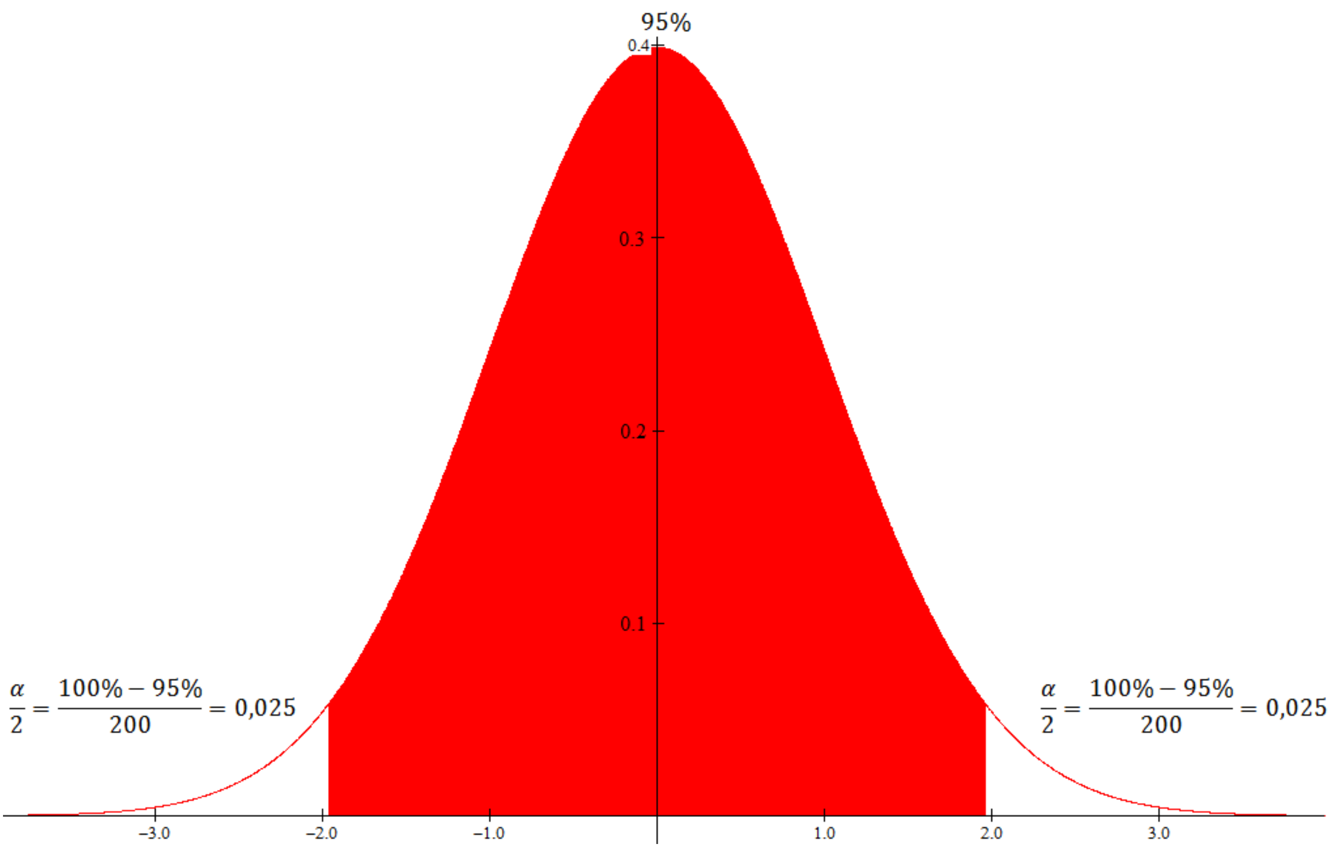
$\frac{\alpha}{2} = \frac{0{,}05}{2} = 0{,}025$

Otra forma para calcular la proporción de la cola superior e inferior de la distribución es aplicando la siguiente fórmula:

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

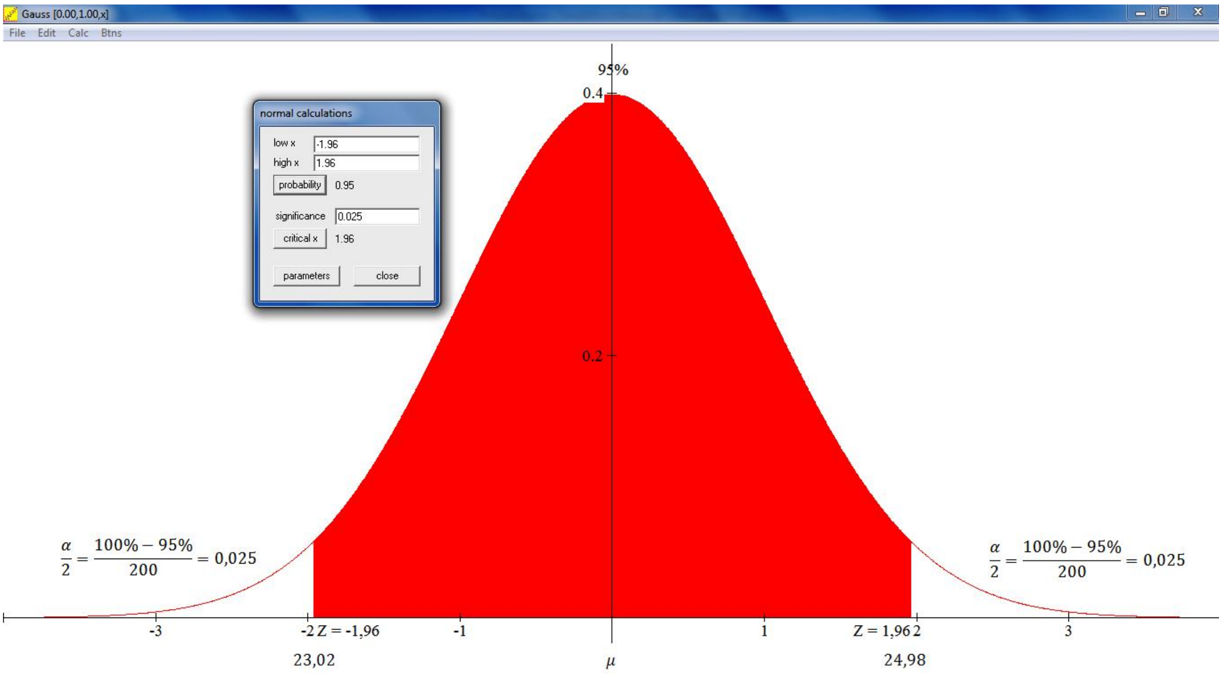
$\frac{\alpha}{2} = \frac{100\% - 95\%}{200} = 0{,}025$

El siguiente gráfico ilustra lo calculado:



6) Si $\bar{X} = 24\,; \sigma = 3\ y\ n = 36$ construya para la media poblacional μ una estimación de intervalo de confianza del 95%

Solución:
Realizando un gráfico ilustrativo en Winstats y Paint se obtiene:



Con lectura en la tabla de la distribución normal para un área de 0,025 se obtiene Z = -1,96. Por simetría se encuentra el otro valor Z = 1,96

Remplazando valores y realizando lo cálculos se obtiene:

$$\bar{X} - Z\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z\frac{\sigma}{\sqrt{n}}$$
$$24 - 1{,}96\frac{3}{\sqrt{36}} \leq \mu \leq 24 + 1{,}96\frac{3}{\sqrt{36}}$$

$$23{,}02 \leq \mu \leq 24{,}98$$

Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F
1	$\bar{X}$	24				
2	σ	3				
3	n	36				
4	Confianza	95				
5	α	0,05	=(100-B4)/100			
6	$Z\frac{\sigma}{\sqrt{n}}$	0,98	=INTERVALO.CONFIANZA.NORM(B5;B2;B3)			
7						
8	23,02	$\leq \mu \leq$	24,98			
9	=B1-B6		=B1+B6			

Interpretación: Existe un 95% de confianza de que la media poblacional se encuentre entre 23,02 y 24,98

7) En un almacén se está haciendo una auditoria para las facturas defectuosas. De 500 facturas de venta se escoge una muestra de 30, de las cuales 5 contienen errores. Construir una estimación del intervalo de confianza del 95%.

Solución:
Los datos del problema son:
$N = 500$
$n = 30$
$X = 5$
$Confianza = 95\%$

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:

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

$$\frac{30}{500} \cdot 100\% = 6\%$$

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:

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

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

$$\frac{\alpha}{2} = \frac{100\% - 95\%}{200} = 0{,}025$$

Con lectura en la tabla de la distribución normal para un área de 0,025 se obtiene Z = -1,96, y por simetría Z =1,96

Calculando la proporción de la muestra se obtiene:

$$p = \frac{X}{n} = \frac{5}{30} = 0{,}167$$

Calculando el intervalo de confianza se obtiene:

$$p - Z\sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}} \leq \pi \leq p + Z\sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$$

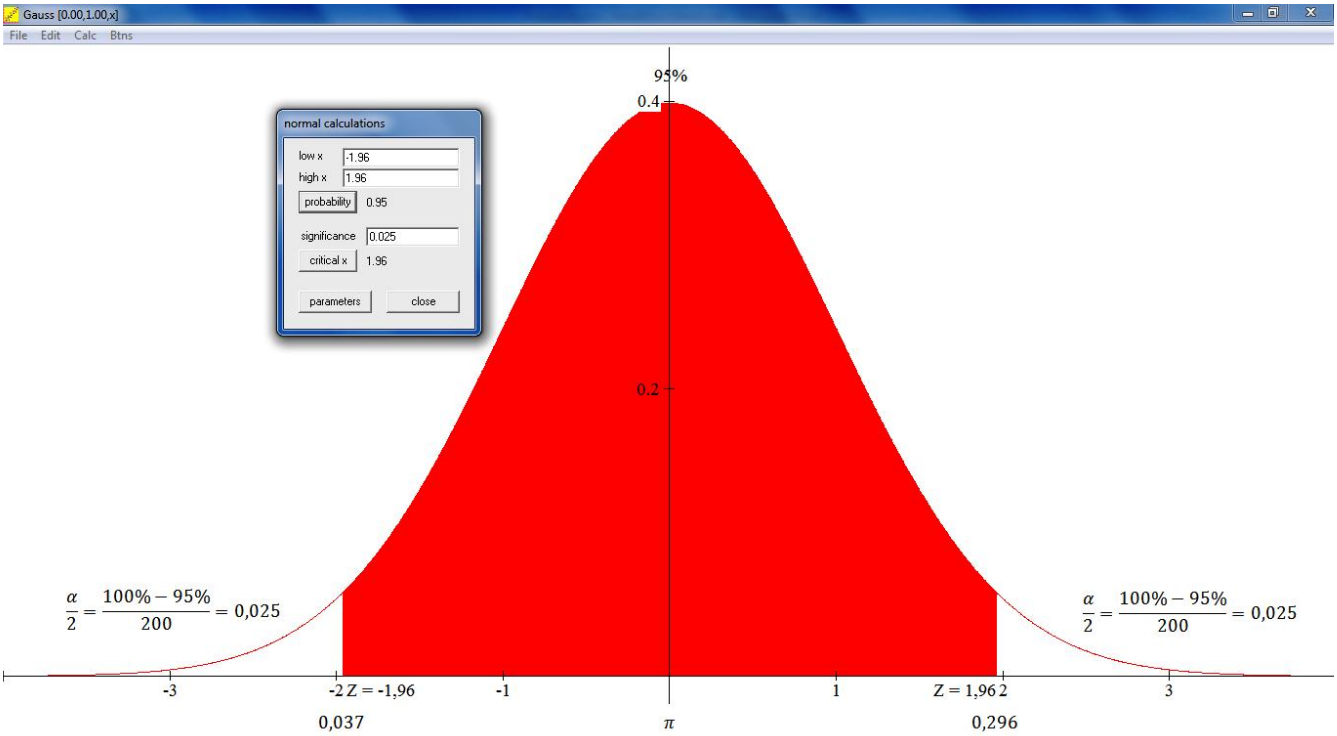
$$0{,}167 - 1{,}96\sqrt{\frac{0{,}167(1-0{,}167)}{30}}\sqrt{\frac{500-30}{500-1}} \leq \pi \leq 0{,}167 + 1{,}96\sqrt{\frac{0{,}167(1-0{,}167)}{30}}\sqrt{\frac{500-30}{500-1}}$$

$$0{,}037 \leq \pi \leq 0{,}296$$

Los cálculos en Excel se muestran en la siguiente figura:

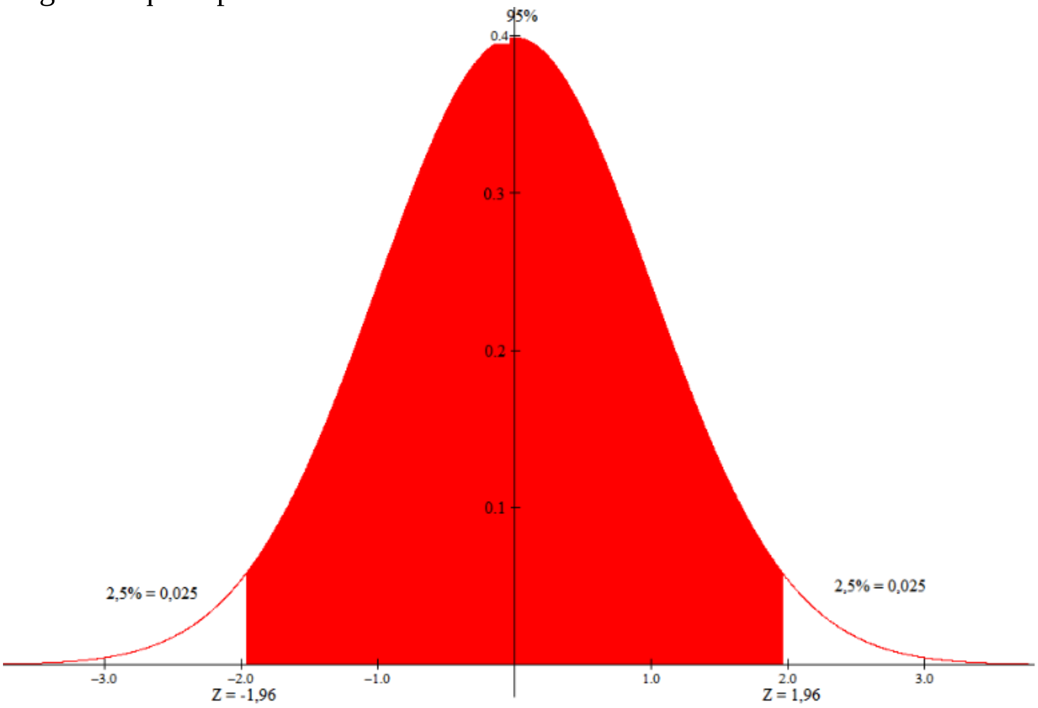
	A	B	C	D	E	F	G	H
1	N	500						
2	n	30						
3	X	5						
4	Confianza	95						
5	$\frac{n}{N}\cdot 100\% > 5\%$	6						
6								
7	$\frac{\alpha}{2}$	0,025						
8								
9	Z	-1,96	=INV.NORM.ESTAND(B7)					
10	Z	1,96	=B9*-1					
11	$p=\frac{X}{n}$	0,167	=B3/B2					
12								
13	$p-Z\sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}\leq \pi \leq p+Z\sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$							
14								
15		0,037	$\leq \mu \leq$	0,296				
16	=B11-B10*RCUAD(B11*(1-B11)/B2)*RCUAD((B1-B2)/(B1-1))			=B11+B10*RCUAD(B11*(1-B11)/B2)*RCUAD((B1-B2)/(B1-1))				

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



8) Calcular el tamaño de la muestra de una población de 500 elementos con un nivel de confianza del 95%

Solución:
Realizando el gráfico que representa el 95% de confianza se obtiene:



Se tiene N=500, para el 95% de confianza Z = 1,96, y como no se tiene los demás valores se tomará $\sigma = 0{,}5$, y e = 0,05.

Reemplazando valores de la fórmula se tiene:

$$n = \frac{N\sigma^2 Z^2}{e^2(N-1) + \sigma^2 Z^2}$$

$$n = \frac{500 \cdot 0{,}5^2 \cdot 1{,}96^2}{0{,}05^2(500-1) + 0{,}5^2 \cdot 1{,}96^2}$$

$$n = \frac{500 \cdot 0{,}5^2 \cdot 1{,}96^2}{0{,}05^2(500-1) + 0{,}5^2 \cdot 1{,}96^2} = 217$$

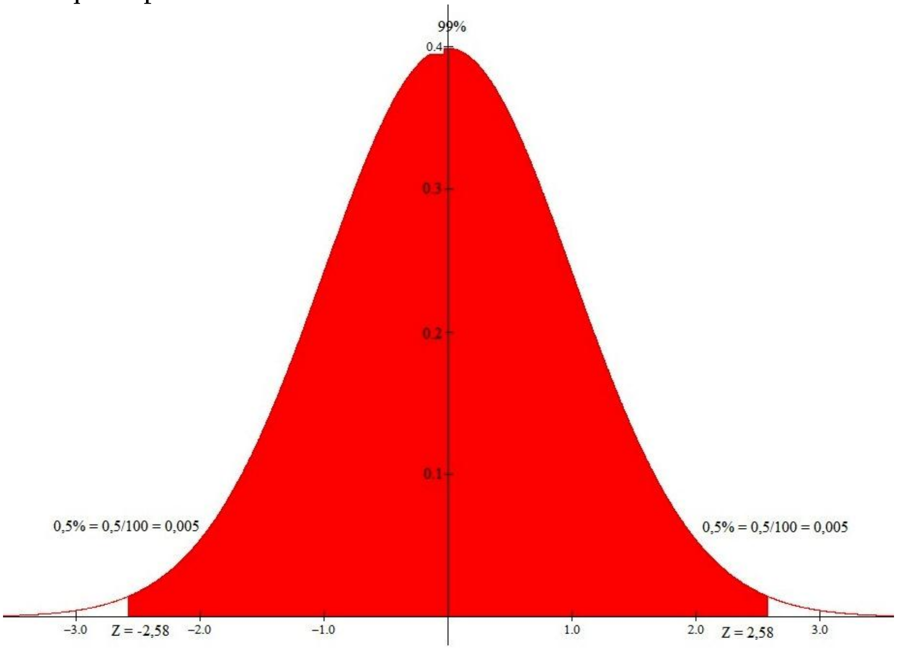
Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E
1	e	0,05			
2	N	500			
3	σ	0,5			
4	Confianza	95			
5	Área a la izquierda de -Z	0,025	=(100-B4)/200		
6	-Z	-1,96	=INV.NORM.ESTAND(B5)		
7	Z	1,96	=-1*B6		
8	$n = \frac{N\sigma^2 Z^2}{(N-1)e^2 + \sigma^2 Z^2}$	217	=B2*B3^2*B7^2/((B2-1)*B1^2+B3^2*B7^2)		
9					

9) Calcular el tamaño de la muestra de una población de 500 elementos con un nivel de confianza del 99%

Solución:

Realizando el gráfico que representa el 99% de confianza se obtiene:



Se tiene N=500, para el 99% de confianza Z = 2,58, y como no se tiene los demás valores se tomará $\sigma = 0{,}5$, y e = 0,05.

Reemplazando valores en la fórmula se obtiene:

$$n = \frac{N\sigma^2 Z^2}{(N-1)e^2 + \sigma^2 Z^2}$$

$$n = \frac{500 \cdot 0{,}5^2 \cdot 2{,}58^2}{(500-1)(\pm 0{,}05)^2 + 0{,}5^2 \cdot 2{,}58^2} = \frac{832{,}05}{2{,}9116} = 285{,}77 = 286$$

Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F
1	e	0,05				
2	N	500				
3	σ	0,5				
4	Confianza	99				
5	Área a la izquierda de -Z	0,005	=(100-B4)/200			
6	-Z	-2,58	=INV.NORM.ESTAND(B5)			
7	Z	2,58	=-1*B6			
8	$n = \frac{N\sigma^2 Z^2}{(N-1)e^2 + \sigma^2 Z^2}$	285,37444	=B2*B3^2*B7^2/((B2-1)*B1^2+B3^2*B7^2)			
9						

10) La duración media de una muestra de 300 focos producidos por una compañía resulta ser de 1620 horas. Se conoce que desviación típica de la población es 150 horas. Comprobar la hipótesis $\mu = 1600$ contra la hipótesis alternativa $\mu \neq 1600$ horas con un nivel de significación de 0,05 si la muestra fue tomada de 5000 focos.

Solución:

Los datos son:
$n = 300$
$\bar{x} = 1620$
$\sigma = 150$
$\alpha = 0{,}05$
$N = 5000$

Las hipótesis son:
$H_0: \mu = 1600$
$H_1: \mu \neq 1600$

Al observar $H_1: \mu \neq 1600$ se trata de una prueba a dos colas, por lo que se tiene que calcular:

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

Como se conoce la desviación estándar de la población σ se debe utilizar la distribución normal. Con lectura en la tabla para un área de 0,025 le corresponde un valor $Z_{tabla} = \pm 1{,}96$. Se toma en cuenta el valor positivo y el negativo porque se trata de una prueba de hipótesis a dos colas.

Como se tiene como dato el tamaño de la población se tiene que verificar si cumple con la condición para utilizar el factor finito de corrección.

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

$$\frac{300}{5000} \cdot 100\% = 6\%$$

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

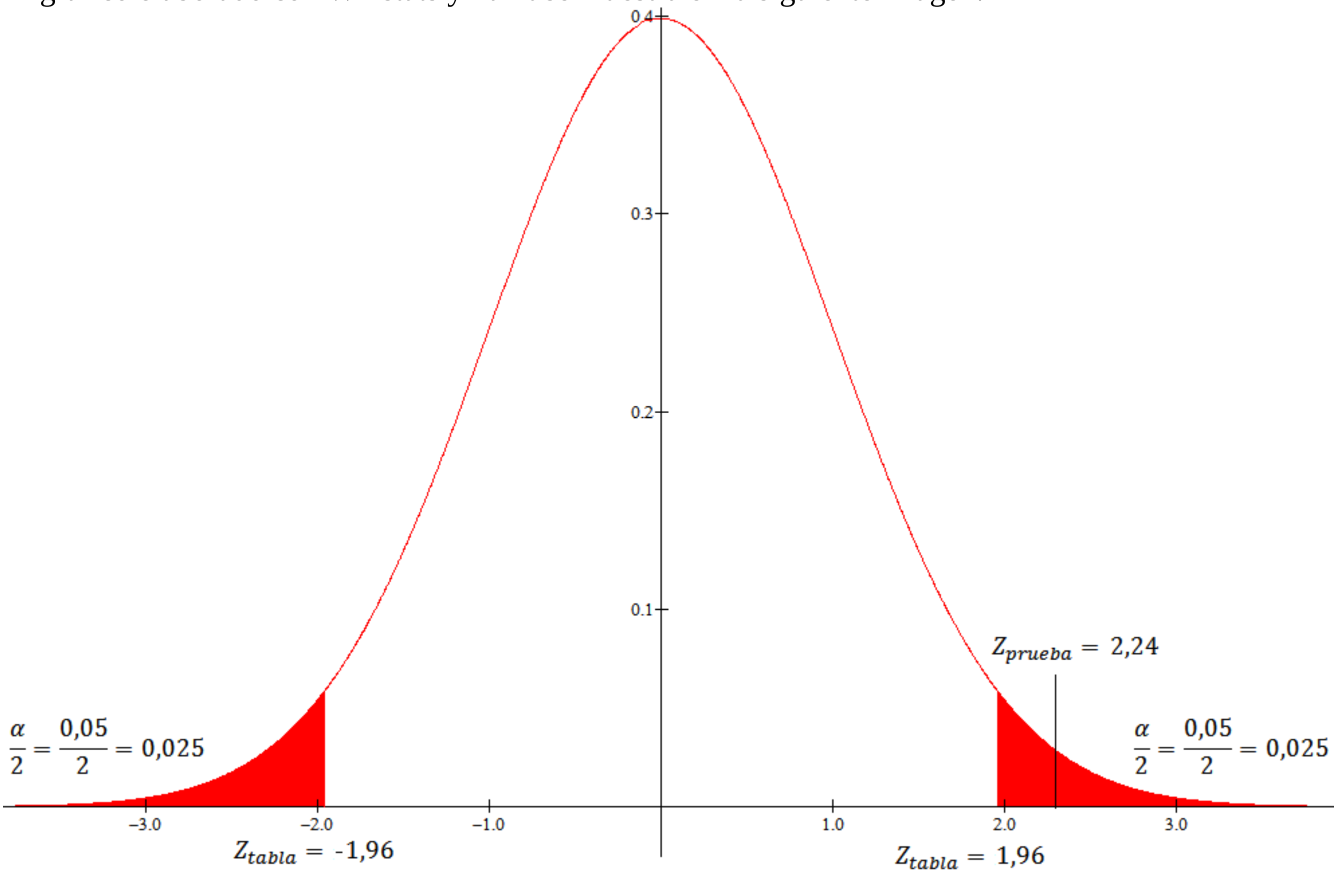
$$Z_{prueba} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}}$$

$$Z_{prueba} = \frac{1620 - 1600}{\frac{150}{\sqrt{300}} \cdot \sqrt{\frac{5000-300}{5000-1}}} = 2{,}24$$

Los cálculos en Excel se muestran en la siguiente imagen:

	A	B	C
1	N	5000	
2	n	300	
3	$\bar{x}$	1620	
4	σ	150	
5	μ	1600	
6	$H_0: \mu = 1600$		
7	$H_1: \mu \neq 1600$		
8	α	0,05	
9	$\frac{\alpha}{2}$	0,025	=B8/2
10			
11	$\frac{n}{N} \cdot 100\% > 5\%$	6,00	=(B2/B1)*100
12			
13	Z_{tabla}	-1,96	=INV.NORM.ESTAND(B9)
14		1,96	=B13*-1
15	$Z_{prueba} = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}}$	2,24	=(B3-B5)/(B4/RCUAD(B2))*RCUAD((B1-B2)/(B1-1))
16			

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



Decisión: Dado que $2,24 > Z_{tabla} \pm 1,96$ se rechaza la H_0, y por lo tanto se acepta H_1

11) La media de las calificaciones de dos muestras de 15 estudiantes de primer semestre en la asignatura de Estadística de la universidad UTN resulta ser de 7 y 8,5. Se sabe que la desviación típica de las calificaciones en esta asignatura fue en el pasado de 1,5. Comprobar la hipótesis $\mu_1 = \mu_2$ contra la hipótesis alternativa $\mu_1 < \mu_2$ con un nivel de significación de 0,025.

Solución:

Los datos son:

$n_1 = n_2 = 15$

$\bar{x}_1 = 7$

$\bar{x}_2 = 8,5$

$\sigma_1 = \sigma_2 = 1,5$

$\alpha = 0,025$

Las hipótesis son:

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 < \mu_2$

Como se conoce la desviación estándar de la población σ se debe utilizar la distribución normal. Con lectura en la tabla para un área de 0,025 le corresponde un valor $Z_{tabla} = -1,96$. Se toma en cuenta el valor negativo porque se trata de una prueba de hipótesis a cola izquierda.

Se remplaza valores en la siguiente fórmula porque se conoce la desviación estándar de la población σ y el tamaño de las muestras son iguales y $n_1 + n_2$ es igual o mayor de 30:

$$Z_{prueba} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

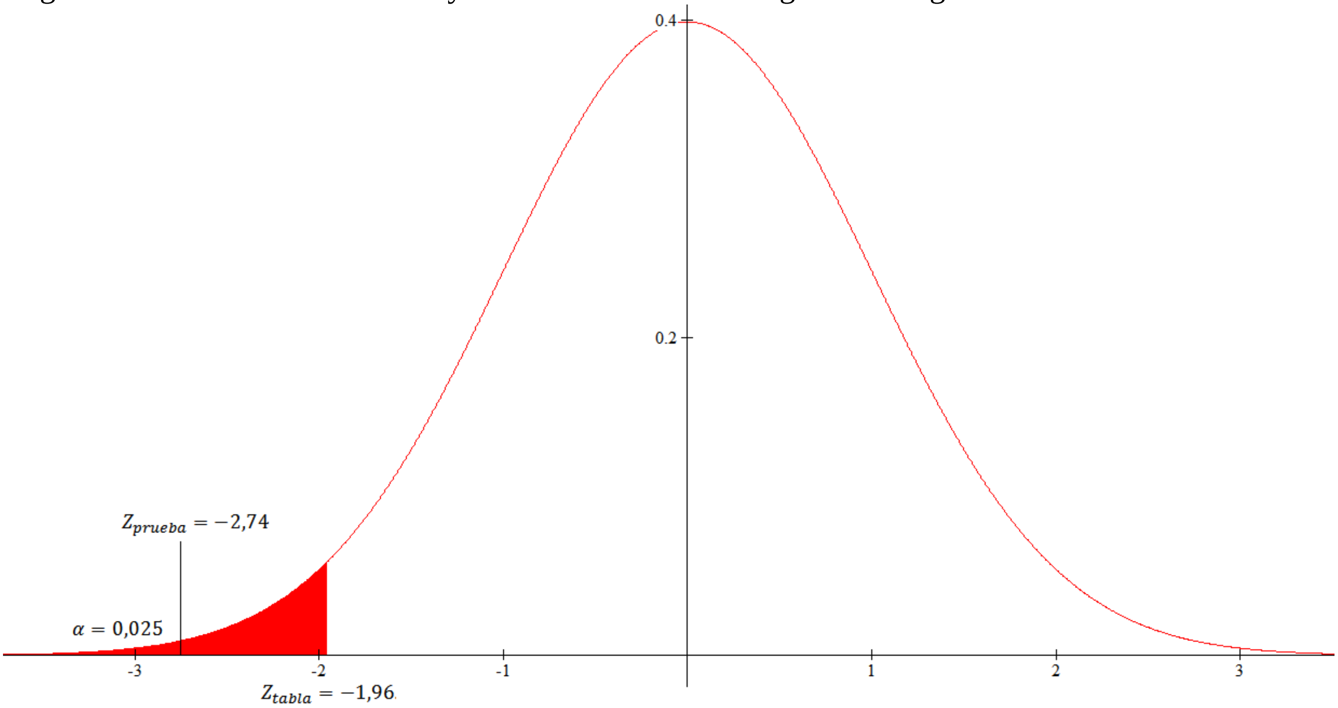
$$Z_{prueba} = \frac{7 - 8,5}{\sqrt{\frac{1,5^2}{15} + \frac{1,5^2}{15}}}$$

$$Z_{prueba} = -2,74$$

Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E
1	n_1	15			
2	n_2	15			
3	$\bar{x}_1$	7			
4	$\bar{x}_2$	8,5			
5	σ_1	1,5			
6	σ_2	1,5			
7	α	0,025			
8	$H_0: \mu_1 = \mu_2$				
9	$H_1: \mu_1 < \mu_2$				
10	Z_{tabla}	-1,96	=INV.NORM.ESTAND(B7)		
11	$Z_{prueba} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	-2,74	=(B3-B4)/RCUAD(B5^2/B1+B6^2/B2)		
12					
13					
14					

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



Decisión: Se rechaza H_0, ya que $Z_{prueba} < Z_{tabla}$

TAREA DE INTERAPRENDIZAJE

1) Realice un organizador gráfico sobre la distribución normal

2) Empleando Excel grafique la campana de la distribución normal calculando previamente los respectivos valores de función de densidad.

3) Calcule el área bajo la curva de distribución normal empleando la tabla de la distribución normal, Excel y Winstats entre:

3.1) Z = 0 y Z = 1,2

0,3849

3.2) Z = 1,2 y Z = 2,2

0,1012

3.3) Z = -2,2 y Z = -1,2

0,1012

3.4) Z = -2 y Z = 1

0,8186

3.5) Z = -1 y Z = 1

0,6827

3.6) Z = -2 y Z = 2

0,9545

3.7) Z = -3 y Z = 3

0,9973

3.8) A la izquierda de Z = 1,54

0,9382

3.9) A la izquierda de Z = -1,54

0,0618

3.10) A la izquierda de Z = -0,6

0,2743

3.11) A la izquierda de Z = 0,6

0,7257

3.12) Plantee y resuelta un ejercicio similar a los anteriores

4) Halle Z si:

4.1) El área entre 0 y Z es 0,3531

±1.05

4.2) El área entre 0 y Z es 0.3972 y Z es positivo

1,266

4.3) El área a la derecha de Z es 0,115

1,20

4.4) El área a izquierda de Z es 0.8621

1,09

4.5) El área a la izquierda de Z = 0,6692

0,44

4.6) El área entre -1 y Z es 0,8186

2

4.7) El área entre -1 y Z es 0,1499

-2,38 y -0,5

4.8) El área entre -0,5 y Z es 0,2313

-1,42 y 0,1

4.9) El área entre 0,5 y Z es 0,1450

0,12 y 0,98

4.10) Plantee y resuelva un ejercicio similar a los anteriores

5) En cierta área, un conductor promedio recorre una distancia de 1200 millas al mes (1,609 Km al mes), con una desviación estándar de 150 millas. Suponga que el número de millas se aproxima mediante una curva normal, encuentre la probabilidad de todos los automovilistas que recorren entre 1200 y 1600 millas por mes. Resuelva empleando la tabla de la distribución normal, Excel y Winstats

49,6%

6) Se sabe que la longitud de cierto mueble se distribuye en forma normal con una media de 84 cm y una desviación estándar 0,4. Calcular el porcentaje de muebles que cumplen con especificaciones 84±1. Resuelva empleando la tabla de la distribución normal, Excel y Winstats

98,76%

7) Las calificaciones que obtienen alumnos universitarios en un examen siguen una distribución normal, siendo la media igual a 7. El 80% de los alumnos obtienen una calificación inferior a 8. Resuelva empleando la tabla de la distribución normal, Excel y Winstats

7.1) Calcule la desviación típica de las calificaciones

1,2

7.2) Se escoge un alumno al azar, calcule el porcentaje de obtener una calificación superior a 9

4,6%

8) El peso de 200 estudiantes varones de cierta universidad es 151 libras, y la desviación típica es 15 libras. Si los pesos están distribuidos normalmente, calcular empleando la tabla de la distribución normal, Excel y Winstats la probabilidad y el número de estudiantes que pesan:

8.1) Más de 128 libras (más de 128,5 libras)

93,32 %; 187 estudiantes

8.2) Menos de 128 libras (menos de 127,5 libras)

5,82 %; 12 estudiantes

8.3) Exactamente 128 libras (Entre 127,5 y 128,5 libras)

0,86 %; 2 estudiantes

8.4) No más de 128 libras (menos de 128,5 libras)

6,68 %; 13 estudiantes

9) La altura de los árbolos de un bosque sigue una distribución normal con una altura media de 17 m. Se selecciona un árbol al azar. La probabilidad de que la altura del árbol seleccionado sea mayor que 24 metros es 6%. Si se selecciona al azar 100 árboles. Emplear la tabla de la distribución normal, Excel y Winstats

9.1) Calcule la desviación típica de las alturas de los árboles

4,5

9.2) Calcular el número esperado de árboles cuyas alturas varían entre 17m y 24m.

44

10) Las calificaciones de los estudiantes de un colegio siguen una distribución normal con una media de 15. Se selecciona un estudiante al azar. La probabilidad de que la calificación del estudiante sea mayor que 17 es del 8%. Resuelva empleando la tabla de la distribución normal, Excel y Winstats

10.1) Calcule la desviación típica de las calificaciones

1,4

10.2) Calcular la probabilidad de que el estudiante seleccionado tenga una calificación menor que 17

0,92

10.3) La probabilidad de que el estudiante tenga una calificación menor que X es 0,08. Calcular el valor de X.

13

10.4) Se selecciona al azar 100 estudiantes. Calcule el número esperado de estudiantes cuyas calificaciones varían entre 1 y 13

8

11) Plantee y resuelva un problema similar al anterior con datos obtenidos de un colegio de su localidad. Emplear la tabla de la distribución normal, Excel y Winstats

12) Las calificaciones que obtienen los alumnos en un examen, cuya máxima nota es 10, siguen una distribución normal, siendo la media igual a 7. El 70% de los alumnos obtienen una calificación inferior a 8,5. Se escoge un alumno al azar, éste alumno tiene la misma probabilidad de obtener una calificación inferior a 9 que una calificación superior a X. Emplear la tabla de la distribución normal, Excel y Winstats

12.1) Calcule la desviación típica de las calificaciones

2,86

12.1) Calcule el porcentaje de obtener una calificación inferior a 9

75,78%

12.2) Calcule el valor de X

5

13) Plantee y resuelva un problema similar al anterior con datos obtenidos en un examen o prueba de cualquier asignatura de su preferencia. Emplear la tabla de la distribución normal, Excel y Winstats

14) ¿El valor Z para el 95% de confianza es?
15) ¿El valor Z para el 99% de confianza es?

Resolver los siguientes ejercicios de manera manual, con Excel para realizar los cálculos y con Winstats para los gráficos

16) Si $\bar{X} = 24{,}2\,; \sigma = 3\ y\ n = 36$ construya para la media poblacional μ una estimación de intervalo de confianza del

16.1) 95%

$23{,}22 \leq \mu \leq 25{,}18$

16.2) 99%

$22{,}91 \leq \mu \leq 25{,}49$

17) 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 y una desviación estándar de 0,02 pulgadas. Se selecciona una muestra de 100 hojas con una media de longitud del papel de 10,998 pulgadas. Calcular la estimación del intervalo de confianza del

17.1) 95%

$10{,}99408 \leq \mu \leq 11{,}00192$

17.3) 60%

$10{,}99632 \leq \mu \leq 10{,}99968$

18) Si n = 200 y X = 50, construya una estimación del intervalo de confianza con lectura en la tabla, Excel y Winstats para la proporción de la población con confianza del

18.1) 95%

$0{,}19 \leq \pi \leq 0{,}31$

18.2) 90%

$0{,}2 \leq \pi \leq 0{,}3$

19) El editor de un periódico desea estimar la proporción de periódicos impresos con algún defecto, tal como borraduras en exceso, disposición errónea de las hojas, páginas faltantes o duplicadas. Se selecciona una muestra aleatoria de 200 periódicos, 35 de ellos contienen algún tipo de defecto. Realice e interprete un intervalo de confianza del 90% para la proporción de periódicos impresos que tienen defectos con lectura en la tabla, Excel y Winstats.

$13{,}1\% \leq \pi \leq 21{,}9\%$

20) Una empresa telefónica desea estimar la proporción de hogares en los que se contrataría una línea telefónica adicional. Se seleccionó una muestra aleatoria de 500 hogares. Los resultados indican que a un costo reducido, 135 de los hogares contratarían una línea telefónica adicional. Construya e interprete una estimación del intervalo de confianza del 99% de la proporción poblacional de hogares que contratarían una línea telefónica adicional con lectura en la tabla, Excel y Winstats.

$21{,}9\% \leq \pi \leq 32{,}0\%$

21) Con lectura en la tabla, Excel y Winstats calcular el tamaño de muestra con un error de muestreo de ±0,05 para N =500, si se desea tener un nivel de confianza en la estimación de la media poblacional del:

21.3.1) 95%

218

21.3.2) 99 %

285

22) La duración media de una muestra de 200 tubos fluorescentes producidos por una compañía resulta ser de 1620 horas. Se sabe que desviación típica de la población es de 100 horas. Comprobar la hipótesis $\mu = 1600$ contra la hipótesis alternativa $\mu \neq 1600$ horas con un nivel de significación de 0,05.

Dado que $2{,}83 > Z_{tabla} \pm 1{,}96$ se rechaza la H_0

23) Cuál sería el resultado del problema anterior si la muestra hubiese sido tomada de una población de 3500 tubos fluorescentes.

Dado que $2{,}75 > Z_{tabla} \pm 1{,}96$ se rechaza la H_0

24) Una empresa de transportes desconfía de la afirmación de que la vida útil promedio de ciertos neumáticos mínimo de 30000 km con una desviación típica de 1500 km. Para verificar se colocan 50 neumáticos en camiones y se obtiene una vida útil promedio de 29400 km. ¿Qué se puede concluir con estos datos si la probabilidad de Error Tipo I es menor a 0,01?

Dado que $-2{,}83 < Z_{tabla} - 2{,}33$ se Rechaza la H_0

Realizar la prueba de hipótesis con lectura en la tabla, Excel (para los cálculos) y el programa Winstats (para las gráficas)

25) H_0: $\mu_1 = \mu_2$, H_1: $\mu_1 > \mu_2$, $\alpha = 0.05$, $\bar{x}_1 = 20$, $\bar{x}_2 = 18$, $\sigma_1 = \sigma_2 = 3$, $n_1 = n_2 = 36$

Se rechaza H_0, ya que $z_{prueba} > z_{tabla}$

26) La media de las edades de dos muestras de 36 estudiantes de primer semestre de una universidad resulta ser de 20 años y 18 años. Se sabe que la desviación típica de todos los estudiantes fue en el pasado de 3 años. Comprobar la hipótesis $\mu_1 = \mu_2$ contra la hipótesis alternativa $\mu_1 > \mu_2$ con un nivel de significación de 0,05

Se rechaza H_0, ya que $z_{prueba} > z_{tabla}$

REFERENCIAS BIBLIOGRÁFICAS

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

SUÁREZ, Mario, (2011), Distribución Normal con Excel, www.monografias.com/trabajos89/