Estadística Inferencial. Resúmen

Métodos y técnicas que permiten inducir el comportamiento de una población.

Muestreo o selección de la muestra:

1. *Aleatorio simple*: se numera la muestra y se elige al azar.
2. *Aleatorio sistemático*: se elige uno al azar y el resto por intervalos.
3. *Aleatorio estratificado*: Se divide en estratos o clases y se elige uno de cada clase.

Teorema central del límite:

Para medias de muestras grandes. $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Probabilidad de que la media de una muestra se encuentre en un intervalo.

Ejemplo: μ = 500 g y σ = 35 g en cajas de 100 unidades. Calcular la probabilidad de que la media de los pesos de las bolsas de un paquete sea menor que 495 g

$$N\left(500, \frac{35}{\sqrt{100}}\right) \quad N(500, 3.5) \quad p(\bar{x} < 495) = p\left(z < \frac{495-500}{3.5}\right) = p(z < -1.43) = p(z > 1.43) = 1 - p(z \leq 1.43) = 0.0764$$

Estimación de parámetros

Nivel de confianza: Probabilidad de que se encuentre en un intervalo de confianza. Para un 95% $1 - \alpha \rightarrow \alpha = 0{,}05$

Nivel de significación: se designa mediante α.

Intervalo de confianza:

Intervalo en el que sabemos que está un parámetro, con un nivel de confianza $\left(\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$

Para un nivel de confianza de 1- α , $\bar{x}$ de media, tamaño n y desviación σ .

Error de estimación admisible: radio del intervalo de confianza

Error máximo de estimación: $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

Intervalo de confianza para una proporción: q = 1-p

Para una $N\left(p, \sqrt{\frac{pq}{n}}\right)$ es: $\left(p' - z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}, p' + z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}\right)$ **error máximo de estimación** es: $E = z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}$

Hipótesis estadísticas

Test estadístico para extraer conclusiones que permitan aceptar o rechazar una hipótesis previamente emitida.

- *Hipótesis nula* H_0. Hipótesis emitida que queremos probar.
- *Hipótesis alternativa* H_1: hipótesis contraria a la nula

Contraste de hipótesis.

A partir de un nivel de confianza β = 1 - α (α nivel de significación de 0,05 ó 0,01)

Proceso: 1º Enunciar la hipótesis nula H_0 y la alternativa H_1.

2º Hallar $z_{\alpha/2}$ bilateral y zona de aceptación del parámetro muestral ($\bar{x}$ o p')

Si el valor del parámetro muestral está dentro de la zona de la aceptación, se acepta la hipótesis con un nivel de significación α. Si no, se rechaza.

Región de aceptación para un intervalo de probabilidad $\bar{x}$ o p':

Contraste bilateral: $\left(\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$ o bien: $\left(p - z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}, p + z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}\right)$

Contraste unilateral: $\left(\mu - z_{\alpha} \cdot \frac{\sigma}{\sqrt{n}}, \infty\right)$ o $\left(-\infty, \mu + z_{\alpha} \cdot \frac{\sigma}{\sqrt{n}}\right)$

Comparación de medias:

- t-Student: Pruebas de muestras cuantitativas:
- Xi-Cuadrado: Pruebas de datos cualitativos.

T-Student: Similar a la normal. Se desconoce la σ. Con n-1 grados de libertad: $t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$

Estadística Inferencial.

La estadística inferencial nos permite estimar características desconocidas como la media de una población o la proporción de la población a partir de muestras.

Existen dos tipos de estimaciones:

- Una estimación puntual: es el valor de un solo estadístico de muestra.
- Una estimación del intervalo de confianza: Rango de números, llamado intervalo, construido alrededor de la estimación puntual.

El intervalo de confianza se construye de manera que la probabilidad del parámetro de la población se localice en algún lugar dentro del intervalo conocido.

Ejemplo: Suponga que quiere estimar la media de todos los alumnos en su universidad.

La media para todos los alumnos es una media desconocida de la población, simbolizada como μ. Usted selecciona una muestra de alumnos, y encuentra que la media es de 5,8. La muestra de la media $\bar{X} = 5{,}8$ es la estimación puntual de la media poblacional μ. ¿Qué tan preciso es el 5,8? Para responder esta pregunta debe construir una estimación del intervalo de confianza.
Recuerde que la media de la muestra $\bar{X}$ es una estimación puntual de la media poblacional μ.

Sin embargo, la media de la muestra puede variar de una muestra a otra porque depende de los elementos seleccionados en la muestra. Tomando en cuenta la variabilidad de muestra a muestra, se aprenderá a desarrollar la estimación del intervalo para la media poblacional.

El intervalo construido tendrá una confianza especificada de la estimación correcta del valor del parámetro poblacional μ . En otras palabras, existe una confianza especificada de que μ se encuentre en algún lugar en el rango de números definidos por el intervalo.

En general, el *nivel de confianza* se simboliza con $(1-\alpha)\cdot 100\%$, donde α es la proporción de las colas de la distribución que están fuera del intervalo de confianza. La proporción de la cola superior e inferior de la distribución es $\alpha/2$

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 (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.

Ejemplo ilustrativo

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

Solución:

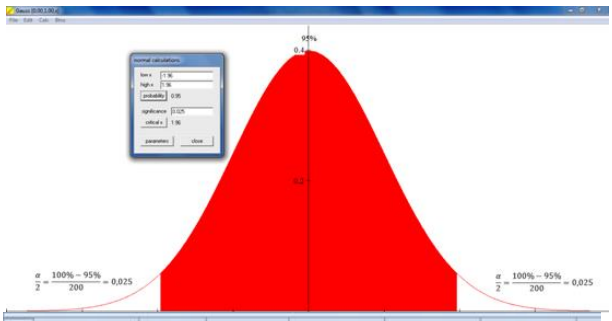
Leyendo en la *tabla de la distribución normal* tenemos que para un área de 0,025 se obtiene Z = -1,96. Por simetría se encuentra el otro valor Z = 1,96
Remplazando valores y realizando lo cálculos se obtiene:

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

$$24 - 1{,}96\frac{3}{\sqrt{36}} \leq \mu \leq 24 + 1{,}96\frac{3}{\sqrt{36}}$$

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

Los cálculos en Excel se muestran en la siguiente figura:
Interpretación: Existe un 95% de confianza de que la media poblacional se encuentre entre 23,02 y 24,98



	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			

Estimación de intervalo de confianza para la media (σ DESCONOCIDA)

Así como la media poblacional μ suele ser desconocida, rara vez se conoce la desviación estándar real de la población σ. Por lo tanto, se requiere desarrollar una estimación del intervalo de confianza de μ usando sólo los estadísticos de muestra $\bar{X}$ y S.

Se emplea la siguiente fórmula:

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

Donde t_{n-1} es el valor crítico de la distribución t con n-1 grados de libertad para un área de $\alpha/2$ en la cola superior

La distribución t supone que la población está distribuida normalmente. Esta suposición es particularmente importante para n < 30. Pero cuando la población es finita y el tamaño de la muestra constituye más del 5% de la población, se debe usar el factor finito de corrección para modificar las desviaciones estándar. Por lo tanto si cumple:

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

Se aplica la ecuación

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

Siendo N el tamaño de la población y n el tamaño de la muestra

Antes de seguir continuando es necesario estudiar la distribución ***t de Student***, 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. Publicado bajo el seudónimo de "Student".

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

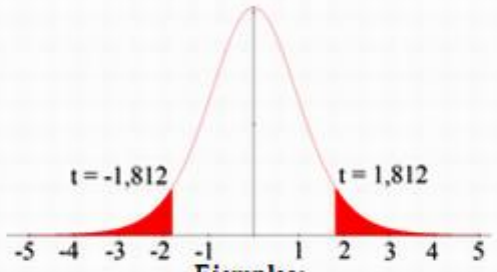
$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

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.

TABLA N° 4
DISTRIBUCIÓN t DE STUDENT



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

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

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:

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).

Ejemplos ilustrativos:

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:

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

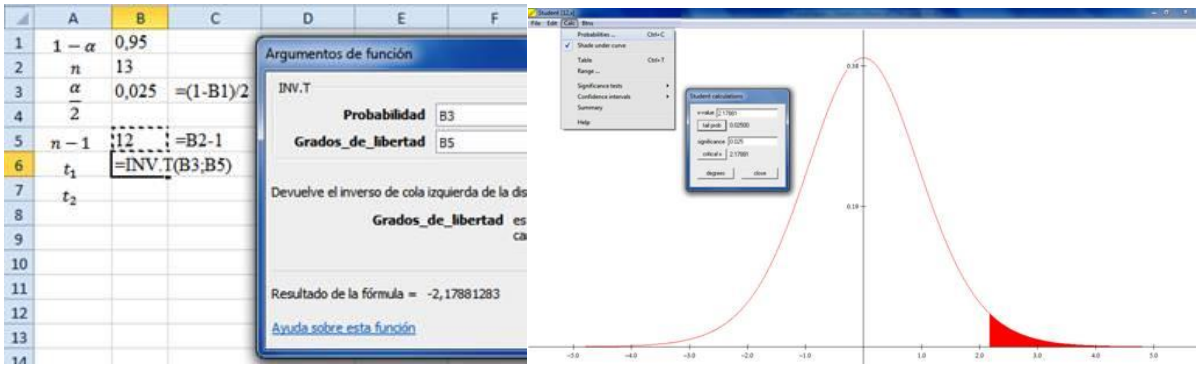
Calculando los grados de libertad se tiene:

$$n-1 = 13-1 = 12$$

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

Los cálculos en Excel y el gráfico se muestran en las siguientes figuras:

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
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



2) *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: Datos del problema:

$\mu = 11$
$N = 500$
$n = 29$
$\bar{X} = 10{,}998$
$S = 0{,}02$
$Confianza = 99\%$

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:

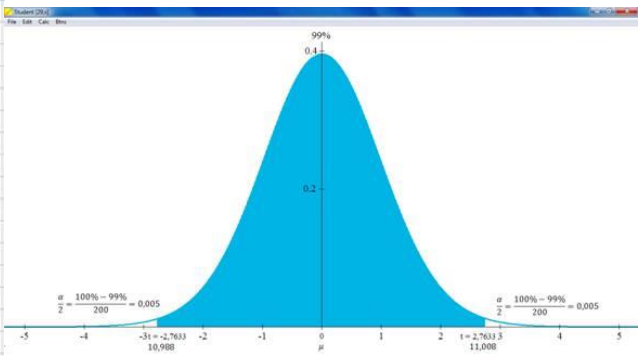
$$\bar{X} - t_{n-1}\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \leq \mu \leq \bar{X} + t_{n-1}\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

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

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

Los cálculos en Excel y el gráfico se muestran en las siguientes figuras:

	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

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

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

Ejemplo ilustrativo

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:

$$Confianza = 95\% \qquad \begin{matrix} N = 500 \\ n = 30 \\ X = 5 \end{matrix}$$

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:

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

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:

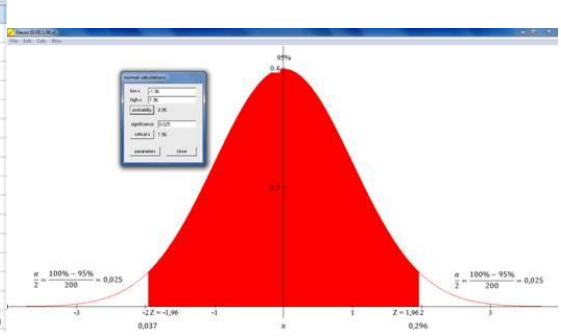
$$p - Z\sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}} \leq \pi \leq p + Z\sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$$

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

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

Los cálculos en Excel y gráficos 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))				



INTERVALO DE CONFIANZA PARA μ; CON σ DESCONOCIDA

Si $\bar{x}$ y s son la media y la desviación estándar de una muestra aleatoria de una población normal con varianza σ^2, desconocida, un intervalo de confianza de $(1-\alpha)100\%$ para μ es:

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

donde $t_{\alpha/2}$ es el valor t con v = n-1 grados de libertad, que deja un área de $\alpha/2$ a la derecha.

Se hace una distinción entre los casos de σ conocida y σ desconocida al calcular las estimaciones del intervalo de confianza. Se debe enfatizar que para el primer caso se utiliza el teorema del límite central, mientras que para σ desconocida se hace uso de la distribución muestral de la variable aleatoria t. Sin embargo, el uso de la distribución t se basa en la premisa de que el muestreo se realiza de una distribución normal. En tanto que la distribución tenga forma aproximada de campana, los intervalos de confianza se pueden calcular cuando la varianza se desconoce mediante el uso de la distribución t y se puede esperar buenos resultados.

Con mucha frecuencia los estadísticos recomiendan que aun cuando la normalidad no se pueda suponer, con σ desconocida y $n \geq 30$, s puede reemplazar a σ y se puede utilizar el intervalo de confianza:

$$\bar{x} \pm z_{\alpha/2}\, s/\sqrt{n}$$

Por lo general éste se denomina como un *intervalo de confianza de muestra grande*. La justificación yace sólo en la presunción de que con una muestra grande como 30, s estará muy cerca de la σ real y de esta manera el teorema del límite central sigue valiendo. Se debe hacer énfasis en que esto es solo una aproximación y que la calidad de este enfoque mejora a medida que el tamaño de la muestra crece más.

Ejemplo:

El contenido de siete contenedores similares de ácido sulfúrico son 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, y 9.6 litros. Encuentre un intervalo de confianza del 95% para la media de todos los contenedores si se supone una distribución aproximadamente normal.

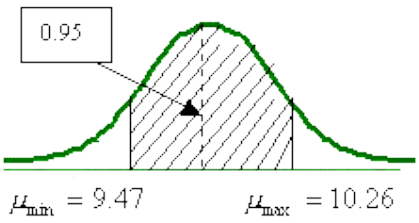
Solución:

La media muestral y la desviación estándar para los datos dados son:

$\bar{x} = 10$ y s= 0.283

En la tabla se encuentra que $t_{0.025}$=2.447 con 6 grados de libertad, de aquí, el intervalo de confianza de 95% para μ es:

$$10.0 - (2.477)\left(\frac{0.283}{\sqrt{7}}\right) < \mu < 10.0 + (2.477)\left(\frac{0.283}{\sqrt{7}}\right)$$



Con un nivel de confianza del 95% se sabe que el promedio del contenido de los contenedores está entre 9.47 y 10.26 litros.