UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE POSGRADO EN CIENCIAS DE LA ADMINISTRACIÓN

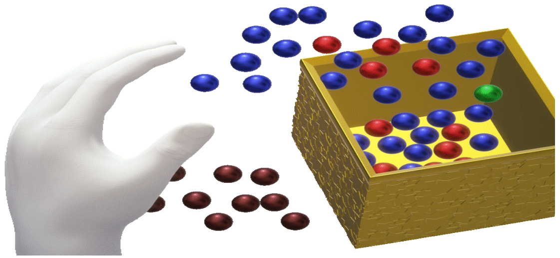
MAESTRÍA EN AUDITORÍA

Métodos Cuantitativos Aplicados a la Auditoría

Dra. María del Rosario Granados Sánchez

marzo – abril 2022






Estimación de parámetros

4.1 Estimación puntual. Medias totales y proporciones
4.2 Estimación por intervalos Medias totales y proporciones
4.3 Trabajo con muestras pequeñas
4.4 Proyección de los errores a la población

La inferencia estadística más simple es la Estimación Puntual, se calcula un valor único (estadístico) con los datos muestrales para estimar un parámetro poblacional.

¿Qué es un estimador?

Es una variable aleatoria que por su naturaleza tiene una distribución (muestral) teórica.

Parámetro poblacional $\Downarrow$ Población μ y σ

Estadístico muestral $\Downarrow$ Muestra $\bar{x}\, y\, s$

$$\boxed{\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}} = \frac{\$\ 1{,}554{,}420}{30} = \textbf{\$ 51, 814}$$

$$\boxed{s = \sqrt{s^2} = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}} = \sqrt{\frac{325009260}{29}} = \textbf{\$ 3,348}$$

$$\boxed{\bar{p} = \frac{x}{n}} = \frac{19}{30} = \textbf{0.63}$$

Parámetro poblacional	Valor del parámetro	Estimador puntual	Estimación puntual
μ	51 800	x	51 814
σ	4 000	s	3 348
p	0.60	p	0.63

Estimación puntual

$$\bar{\chi} = \mu$$
$$s = \sigma$$
$$\tilde{\rho} = \rho$$

Estadístico muestral ➔ Inferencia sobre un parámetro poblacional.

Distribución muestral o de muestreo

$$\bar{\chi} = 51,814$$

$$\tilde{\rho} = 0.63$$

500 muestras aleatorias:

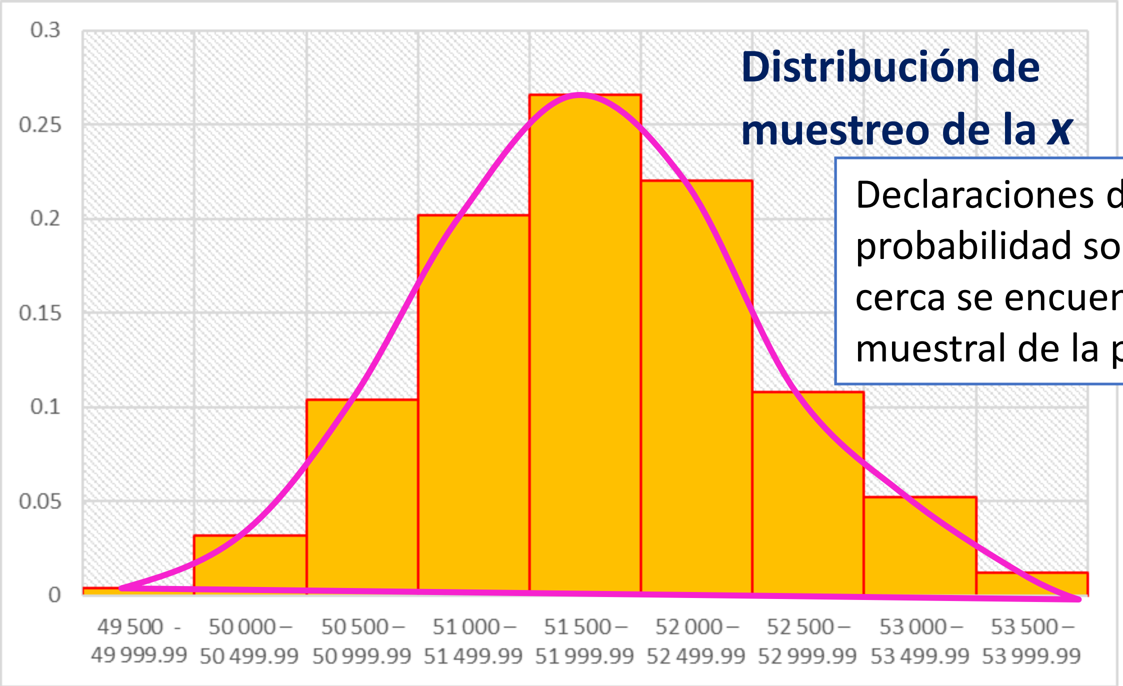
No. Muestra	*x*	*p*
1	51 814	0.63
2	52 670	0.70
…	…	…
500	51 752	0.67

Estimaciones puntuales diferentes

¿Cómo sabemos con qué frecuencia se presentan estás distribuciones?

Segmentamos los salarios y construimos la tabla de frecuencias

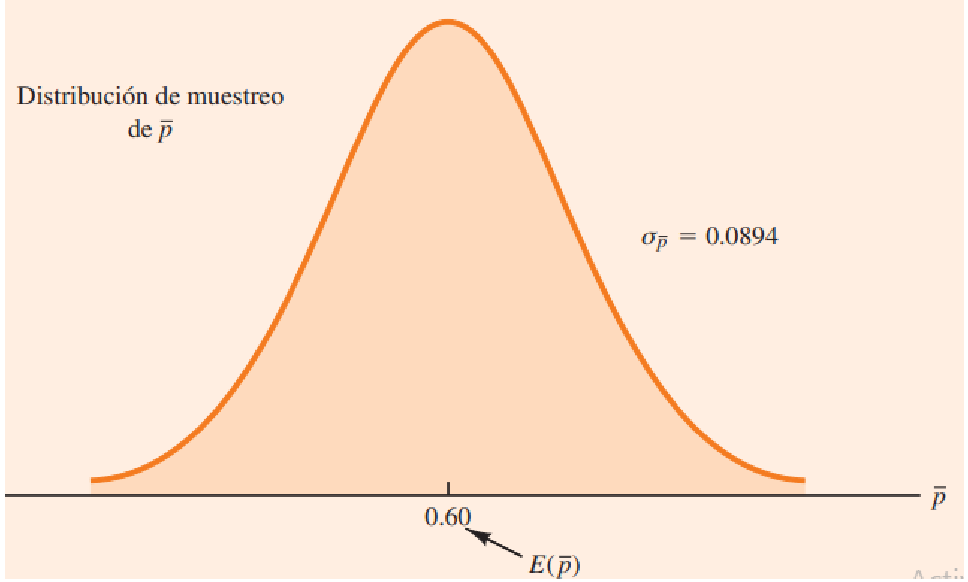
Sueldo medio mensual	Frecuencia	Frecuencia relativa
49 500 - 49 999.99	2	0.004
50 000 – 50 499.99	16	0.032
50 500 – 50 999.99	52	0.104
51 000 – 51 499.99	101	0.202
51 500 – 51 999.99	133	0.266
52 000 – 52 499.99	110	0.220
52 500 – 52 999.99	54	0.108
53 000 – 53 499.99	26	0.052
53 500 – 53 999.99	6	0.012
Σ	**500**	**1**



Declaraciones de probabilidad sobre qué tan cerca se encuentra la media muestral de la poblacional.

Distribución de muestreo de la proporción muestral

Cada distribución de muestreo de $\bar{p}$ de las 500 muestras de tamaño “n”:



A la distribución de muestreo de cualquier estadístico determinado se le llama: distribución de muestreo del estadístico

Características de distribución de muestreo de la media ($\bar{x}$):

La distribución muestral de ($\tilde{x}$) es la distribución de probabilidad de todos los posibles valores de la media muestral ($\tilde{x}$)

a. Valor esperado

Generar distintas muestras ➔ distintos valores de la $\overline{x}$ ∴ la media de la variable aleatoria $\overline{x}$ es el valor esperado de la $\overline{x}$

$$E(\bar{x}) = \mu$$

Estimador puntual insesgado

b. Desviación estándar de la $(\bar{x})$

$$\sigma_{\chi} = DS\ de\ \bar{x}$$

$$\sigma \quad = DS\ de\ la\ población$$

Si n/N > 0.05
población finita

La DS de la media dependerá de que la población sea finita o infinita.

Población finita

$$\sigma_{\bar{x}} = \underbrace{\sqrt{\frac{N-n}{N-1}}}_{\text{Factor de corrección} \approx 1}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Factor de corrección ≈ 1

La diferencia entre el valor de las DS de $\bar{x}$ para las poblaciones finitas e infinitas se vuelve irrelevante

Población infinita

Error estándar de la media

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Población infinita o finita, pero con un tamaño de muestra ≤ 5% del tamaño de la población.

La DS poblacional es una buena aproximación de la DS muestral.

Estimador puntual:

Propiedades:

1. Insesgadez
2. Eficiencia
3. Consistencia

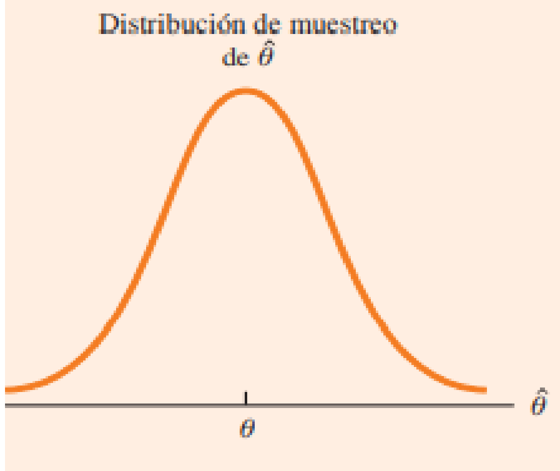
θ = parámetro poblacional de interés

$\hat{\theta}$ = estimador puntual de θ

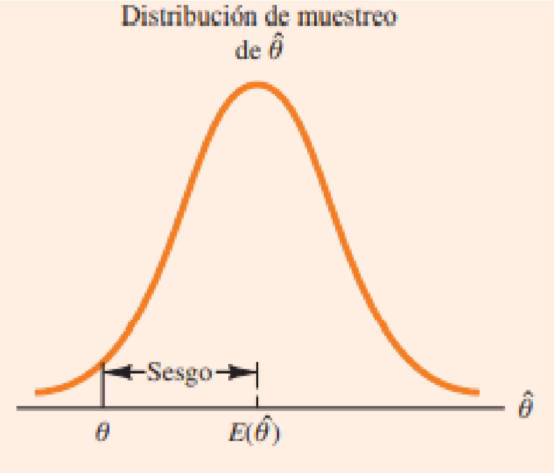
Insesgadez

El estadístico muestral $\hat{\theta}$ es un estimador insesgado del parámetro poblacional, sí:

E $(\hat{\theta}) = \theta$



El parámetro θ se localiza en la media de la distribución de muestreo; $E(\hat{\theta}) = \theta$

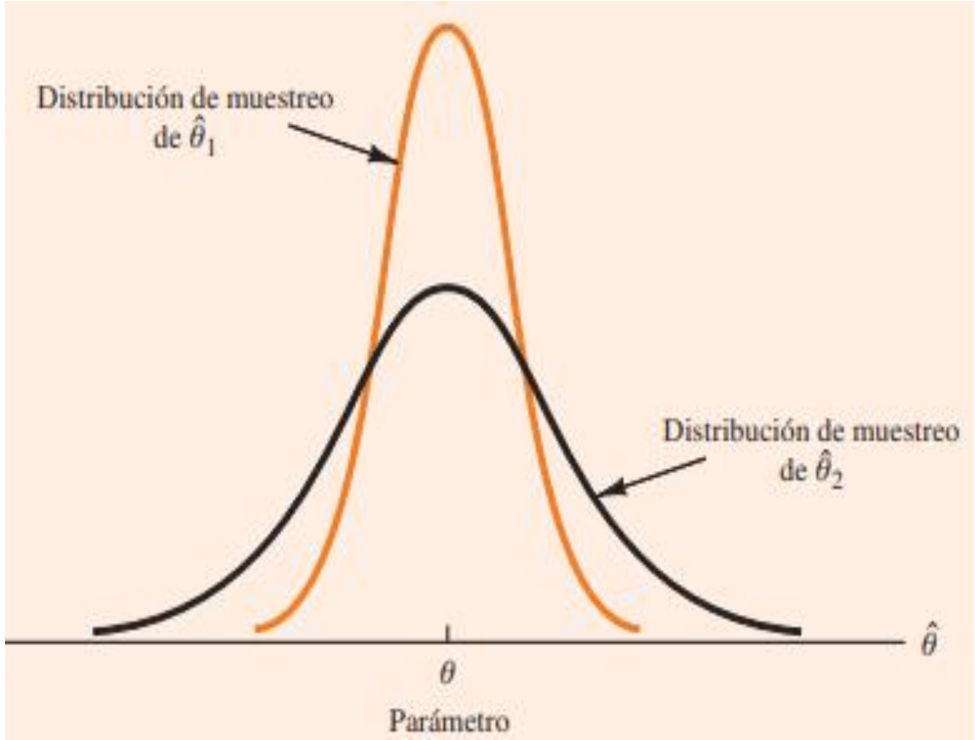


El parámetro θ no se localiza en la media de la distribución de muestreo; $E(\hat{\theta}) \neq \theta$

Estimador puntual:

Eficiencia

Cuando hay un estimador puntual con el menor error estándar, ya que tenderá a proporcionar estimaciones más cercanas al parámetro poblacional.



$\widehat{\theta_1} < \widehat{\theta_2}$

Los valores de $\widehat{\theta_1}$ tienden más posibilidades de estar cerca de θ

1 es relativamente más eficiente que 2

Consistencia

Un estimador puntual es consistente si su valor tiende a estar más cerca del parámetro poblacional a medida que aumenta el tamaño de la muestra.

“n” grande tiende a proporcionar mejor estimación puntual que una pequeña.

Un estimador puntual nos da información sobre el parámetro poblacional, pero no nos dice que tan cerca estamos del valor real.

Necesario estimar un **intervalo de confianza**.

Conjunto de valores obtenidos a partir de los datos muestrales, en el que hay una determinada **probabilidad** de que se encuentre el parámetro.

Estimación por intervalo para la μ y p

$$\hat{\theta}_1 < \theta < \hat{\theta}_2$$

$\bar{x} \pm Margen\ de\ error$

$\bar{p} \pm Margen\ de\ error$

Las distribuciones de muestreo son clave para calcular las estimaciones por intervalo

Situación 1: Media poblacional y varianza conocida

Cuando tenemos una gran cantidad de datos históricos tenemos una varianza conocida (σ).

Recuerden que…la distribución de muestreo de $\bar{x}$ sirve para calcular la probabilidad de que la $\bar{x}$ esté dentro de una distancia dada de μ

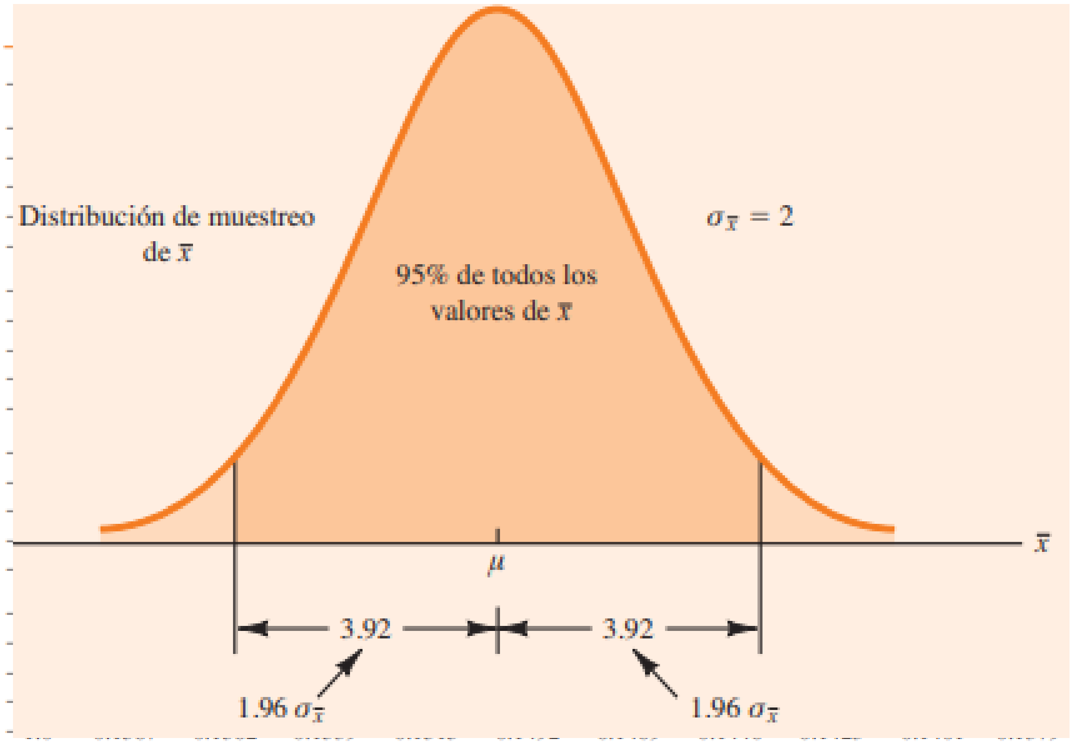
Ejemplo: Tienda departamental selecciona una m.a.s. cada semana de 100 clientes para ver la cantidad que gastan en cada visita a la tienda.

Tenemos:

σ = \$20 $\bar{x}$ = \$82 n=100 ➔ distribución normal

Error estándar =

Tabla de probabilidad normal estándar vamos a encontrar que el 95% de los valores de cualquier variable aleatoria distribuida normalmente aparecen dentro de $\pm$ 1.96 desviaciones estándar de la media (μ)



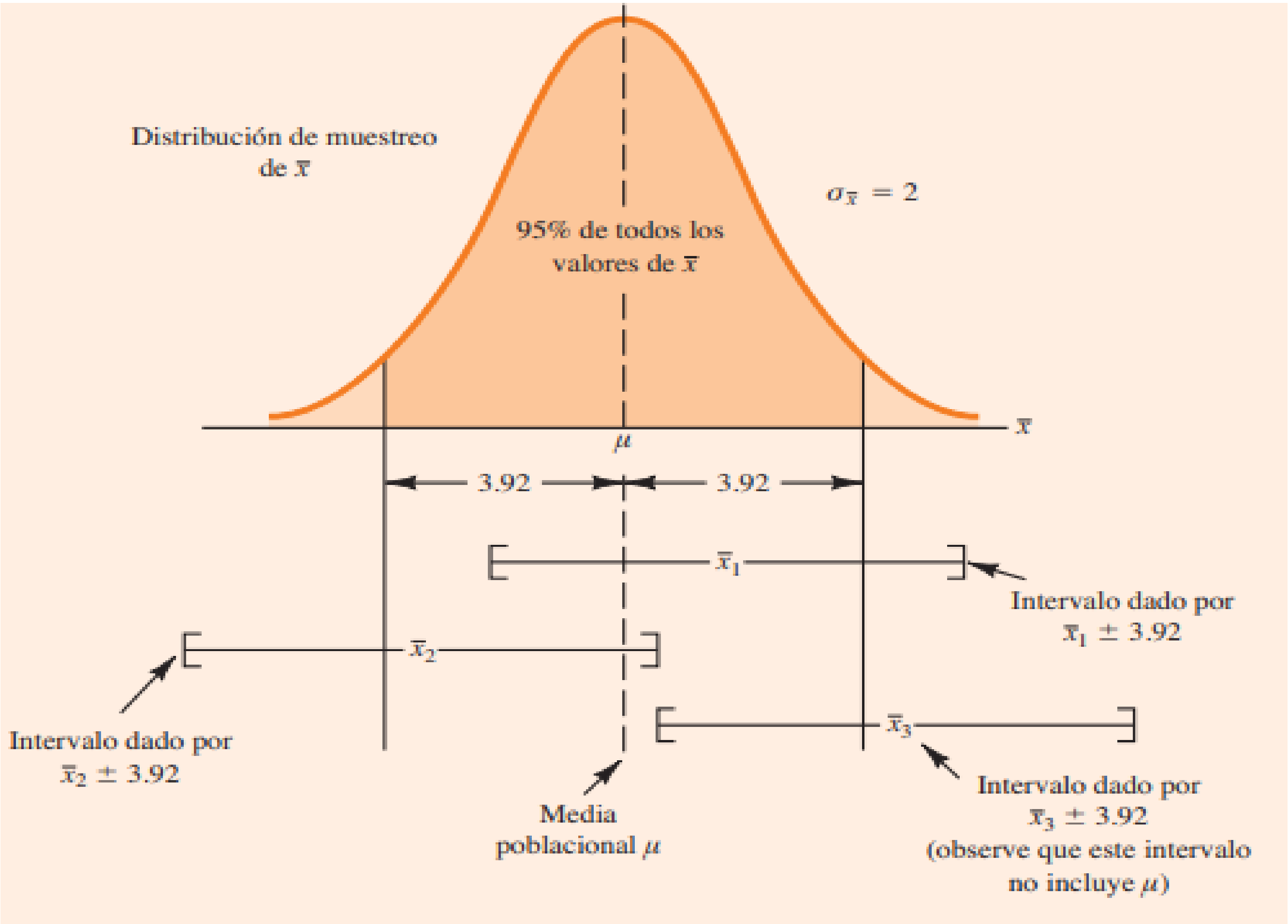
La distribución de muestreo de $\bar{x}$ se distribuye normalmente con una $\sigma_{\bar{x}}$ = 2

El margen de error será = $\pm 1.96\ (\sigma_{\bar{x}})$

El margen de error será = $\pm 1.96\ (2) = 3.92$

Supongamos que tenemos 3 m.a.s diferentes de tamaño n=100

Caso 1: La $\overline{x_1}$
Caso 2: La $\overline{x_2}$
Caso 3: La $\overline{x_3}$



$\bar{x} \pm Margen\ de\ error$

$82 + 3.92 =$ 85.92 Intervalo superior

$82 - 3.92 =$78.08 Intervalo inferior

$Con\ un\ 95\%\ de\ nivel\ de\ confianza, el\ intervalo\ 78.08\ a\ 85.92\ contiene\ a\ la$ μ

95% = 0.95 ➔ Coeficiente de confianza

78.08 a 8.92 ➔ Intervalo de confianza

Estimación por Intervalo

$$\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

En el ejemplo, el coeficiente de confianza estuvo dado por (1-0.05) = 0.95

α = 0.05
α/2 = 0.05 / 2 = 0.025

Buscamos en tablas
$Z_{0.025}$= 1.96

Calcular el intervalo de confianza con un nivel de confianza del 90% y 99%

Resuelva lo siguiente:

1. n=50 σ=6 $\bar{x} = 32$
a. Proporcione un intervalo de confianza de 90, 95 y 99% para la μ

2. Para la media poblacional el intervalo de confianza de 95% resultó de 152 a 160. Si σ=15, ¿cuál es el tamaño de la muestra utilizado?

$$\mathrm{E} = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Despejamos $\sqrt{n}$

$$\sqrt{n} = \frac{Z_{\alpha/2}\sigma}{\mathrm{E}}$$

Elevamos al cuadrado ambos lados de la expresión

$$\boxed{n = \frac{\left(Z_{\alpha/2}\right)^2 \sigma^2}{E^2}}$$

Resumen de los procedimientos para la estimación por intervalo de la media poblacional.

¿Se conoce la σ?

SÍ

Utilice la distribución Z

$$\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

NO

Utilice la "s" para estimar σ

$$\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

Es normal la población

NO

n ≥ 30

NO

Prueba no parámetrica

SÍ

Distribución Z

SÍ

Situación 2: Media poblacional y varianza desconocida

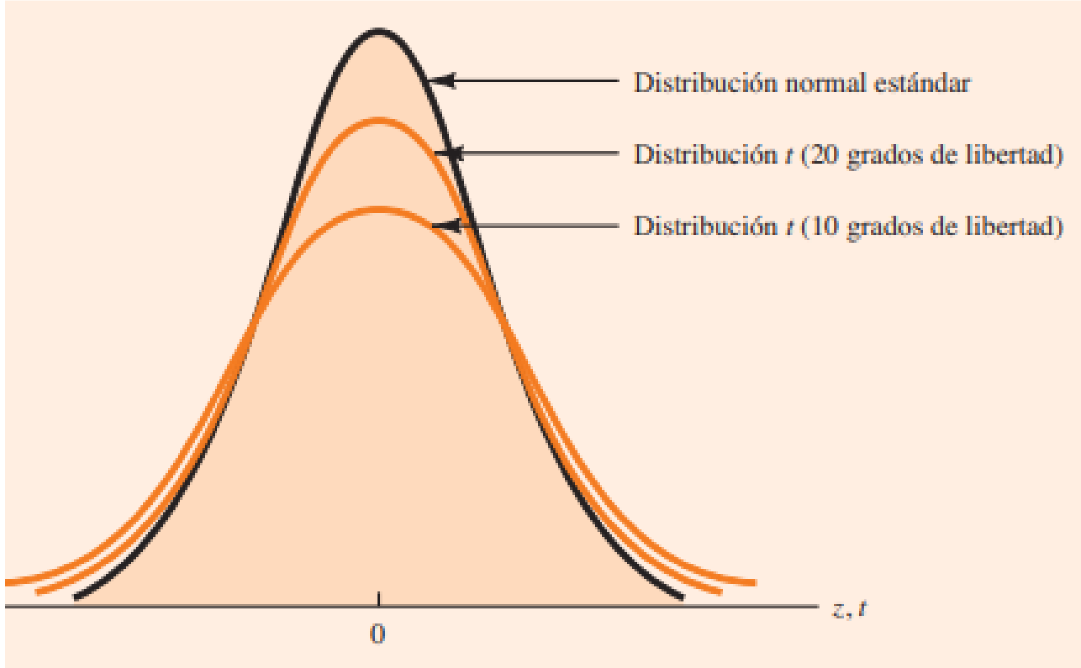
Al utilizar la "s" para estimar la σ, el margen de error de la estimación por intervalo de μ se basan en una distribución de probabilidad conocida como **distribución "t"**

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

Es una familia de distribuciones de probabilidad similar, cada probabilidad depende de un parámetro conocido como **grados de libertad (gl).**

A medida que **aumentan** los **gl** la **diferencia entre** la distribución **normal estándar** y la distribución **t, se reduce.**

Se aplica en situaciones en que la población se desvía significativamente de la normal.

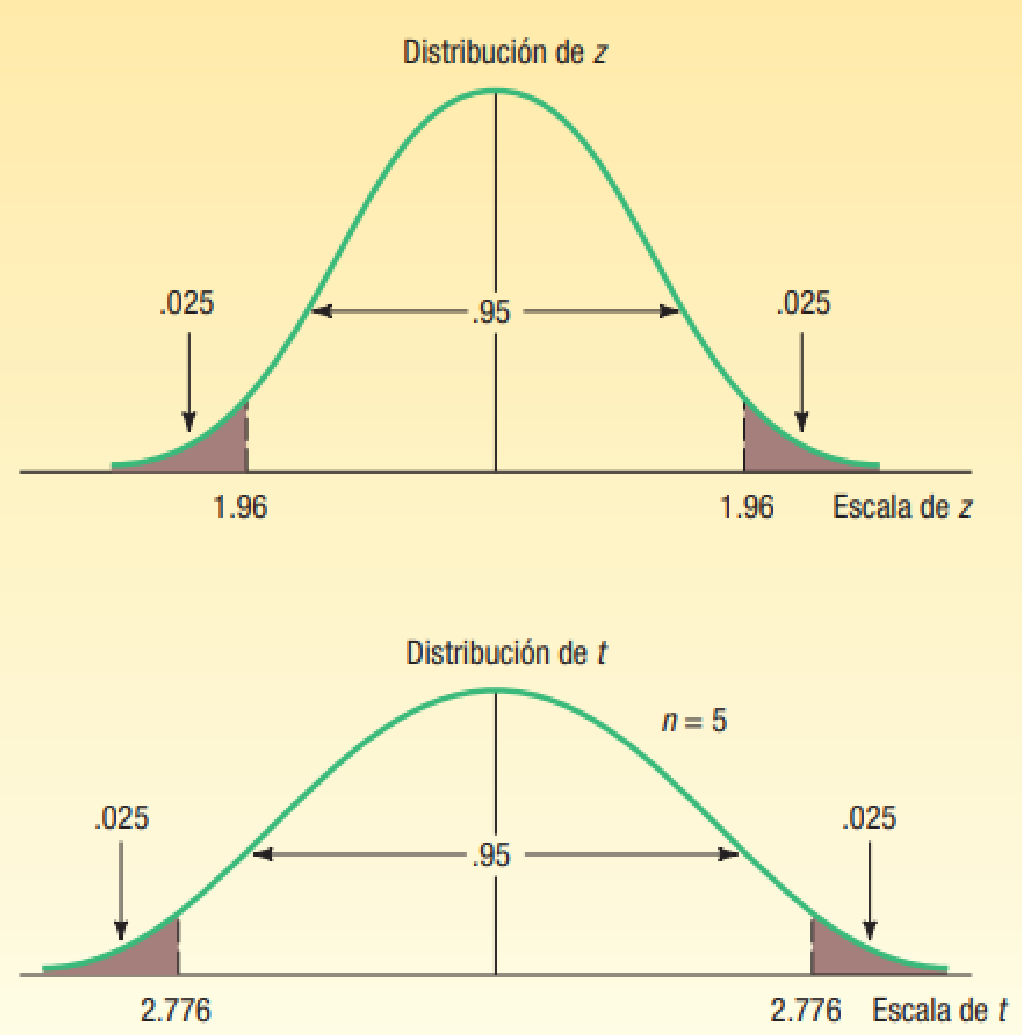


Propiedades de la distribución "t"

Se basan en el supuesto de que la población de interés es de naturaleza normal, o casi normal.

- Como en el caso de la distribución z, es una **distribución continua**.
- Como en el caso de la distribución z, tiene **forma de campana y es simétrica**.
- No existe una distribución t, sino **una familia de distribuciones t**. Todas las distribuciones t **tienen una media de 0**, y sus **desviaciones estándares difieren de acuerdo con el tamaño de la muestra, n**. Existe una distribución t para un tamaño de muestra de 20, otro para un tamaño de muestra de 22, etc. La desviación estándar de una distribución t con 5 observaciones es mayor que en el caso de una distribución t con 20 observaciones.
- **La distribución t se extiende más y es más plana por el centro que la distribución normal estándar.** Sin embargo, conforme se incrementa el tamaño de la muestra, la distribución t se aproxima a la distribución normal estándar, pues los errores que se cometen al utilizar s para estimar disminuyen con muestras más grandes.

Dado que la distribución t de Student posee mayor dispersión que la distribución z, el valor de t en un nivel de confianza dado tiene una magnitud mayor que el valor z correspondiente.



Manejo de la tabla de distribución “t”:

Elegimos 15gl, con un nivel de significancia del 95%.

$t_{\alpha/2}$ **= 0.05/2 = 0.025**

Grados de libertad	Área en la cola superior 0.20	0.10	0.05	0.025	0.01	0.005
1	1.376	3.078	6.314	12.706	31.821	63.656
2	1.061	1.886	2.920	4.303	6.965	9.925
3	0.978	1.638	2.353	3.182	4.541	5.841
4	0.941	1.533	2.132	2.776	3.747	4.604
5	0.920	1.476	2.015	2.571	3.365	4.032
6	0.906	1.440	1.943	2.447	3.143	3.707
7	0.896	1.415	1.895	2.365	2.998	3.499
8	0.889	1.397	1.860	2.306	2.896	3.355
9	0.883	1.383	1.833	2.262	2.821	3.250
10	0.879	1.372	1.812	2.228	2.764	3.169
11	0.876	1.363	1.796	2.201	2.718	3.106
12	0.873	1.356	1.782	2.179	2.681	3.055
13	0.870	1.350	1.771	2.160	2.650	3.012
14	0.868	1.345	1.761	2.145	2.624	2.977
15	0.866	1.341	1.753	2.131	2.602	2.947
16	0.865	1.337	1.746	2.120	2.583	2.921
17	0.863	1.333	1.740	2.110	2.567	2.898
18	0.862	1.330	1.734	2.101	2.552	2.878
19	0.861	1.328	1.729	2.093	2.539	2.861
20	0.860	1.325	1.725	2.086	2.528	2.845
21	0.859	1.323	1.721	2.080	2.518	2.831
22	0.858	1.321	1.717	2.074	2.508	2.819
23	0.858	1.319	1.714	2.069	2.500	2.807
24	0.857	1.318	1.711	2.064	2.492	2.797

A medida que aumentan los gl

$t_{0.025}$ **= 2.131** se aproxima a

$Z_{0.025}$ **= 1.96**

Para más de 100gl, el valor Z de la distribución normal estándar proporciona una buena aproximación del valor “t”

∞ **0.025 = 1.96**

Margen de error y estimación por intervalo

Usamos la "s" para estimar σ:

Sustituimos: $Z_{\frac{\alpha}{2}} \dfrac{\sigma}{\sqrt{n}}$ Por: $t_{\frac{\alpha}{2}} \dfrac{\sigma}{\sqrt{n}}$ **MARGEN DE ERROR**

Podemos realizar la estimación por intervalo de la μ con σ desconocida: $\bar{x} \pm t_{\frac{\alpha}{2}} \dfrac{s}{\sqrt{n}}$

Valor de "t" que proporciona un área de α/2 en la cola superior de la distribución **"t" con n-1 gl**

Ejemplo:
Un banco necesita estimar la media del adeudo de las tarjetas de crédito de la población de las familias de la CDMX.

n= 70 σ= no la conocemos $\bar{x} = \$9{,}312$ $s = \$4{,}007$
Nivel de confianza 95%

1. Buscar en tablas "t Student" el valor al 95%

2. Calculamos la estimación por intervalo

Esto resulta exacto cuando la población tiene una distribución normal. Si no la tiene, el intervalo de confianza será aproximado.

ESTIMACIÓN DE INTER

USO DE UNA MUESTRA PEQUEÑA

Cuando n < 30 y no conocemos la σ:

Ej. Tenemos una muestra de 20 empleados y los días que duró la capacitación de cada uno.

¿Cómo es la distribución muestral?

No es posible concluir que la población sea normal.
Lo correcto es sustituir la distribución normal por "t"

n= 20 σ= no la conocemos $\bar{x} = 51.5$ $s = 6.84$
Nivel de confianza 95%

1. Buscar en tablas "t Student" el valor al 95%

2. Calculamos la estimación por intervalo

$$\boxed{\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}}$$

DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA CUANDO NO CONOCEMOS LA σ

1. Utilizamos un valor planeado de la s
2. Calculada a partir de estudios históricos
3. Realizamos un estudio piloto seleccionando una muestra preliminar
4. Utilizamos el juicio personal para adivinar el mejor valor de σ
   Valor mayor y menor: la diferencia entre los valores proporciona una estimación del rango de datos / 4➔ aproximación a la σ

Intervalo de confianza de una proporción

Fracción, razón o porcentaje que indica la parte de la muestra de la población que posee **un rasgo de interés particular (variable nominal).**

PROPORCIÓN MUESTRAL $$p = \frac{X}{n}$$

Para crear el intervalo de confianza de una proporción, es necesario cumplir con los siguientes supuestos:

1. Las condiciones binomiales han quedado satisfechas:
   a) Los datos de la muestra son resultado de conteos.
   b) Sólo hay dos posibles resultados (lo normal es referirse a uno de los resultados como éxito y al otro como fracaso).
   c) La probabilidad de un éxito permanece igual de una prueba a la siguiente.
   d) Las pruebas son independientes. Esto significa que el resultado de la prueba no influye en el resultado de otra.
2. Los valores $n\pi$ y $n(1-\pi)$ deben ser mayores o iguales que 5. Esta condición permite recurrir al teorema central del límite y emplear la distribución normal estándar, es decir, z, para completar un intervalo de confianza.

El desarrollo del estimador puntual de la proporción de la población y el intervalo de confianza de una proporción de población es similar a hacerlo para una media.

$$p \pm z\sqrt{\frac{p(1-p)}{n}}$$

El sindicato que representa a Bottle Blowers of America (BBA) considera la propuesta de fusión con Teamsters Union. De acuerdo con el reglamento del sindicato de BBA, por lo menos tres cuartas partes de los miembros del sindicato deben aprobar cualquier fusión. Una muestra aleatoria de 2 000 miembros actuales de BBA revela que 1 600 planean votar por la propuesta. ¿Qué es el estimador de la proporción poblacional? Determine el intervalo de confianza de 95% de la proporción poblacional. Fundamente su decisión en esta información de la muestra: ¿puede concluir que la proporción necesaria de miembros del BBA favorece la fusión? ¿Por qué?

1. Calcular la proporción: $p = \frac{X}{n} =$

Se estima que % de la población favorece la propuesta de fusión.

2. Determinar el intervalo de confianza (95%):

$$p \pm z\sqrt{\frac{p(1-p)}{n}} =$$

I.C= ($< p <$)

, por lo tanto, es probable que se apruebe la propuesta de fusión entre los sindicatos

Recordatorio de la interpretación del intervalo de confianza: si la encuesta fue aplicada 100 veces con 100 muestras distintas, los intervalos de confianza construidos a partir de 95 de las muestras contendrán la verdadera proporción de la población.