Estimación con intervalos de confianza

Universidad de Medellín
Ciencia y Libertad

Debido a que las poblaciones por lo general son muy grandes para ser estudiadas en su totalidad, se seleccionan muestras las cuales se pueden utilizar para hacer inferencias sobre las poblaciones. Hay por lo menos dos tipos de estimadores que se utilizan para este propósito:

•El estimador puntual utiliza un estadístico para estimar el parámetro en un solo valor o punto.

•Una estimación por intervalo especifica el rango dentro del cual está el parámetro desconocido.

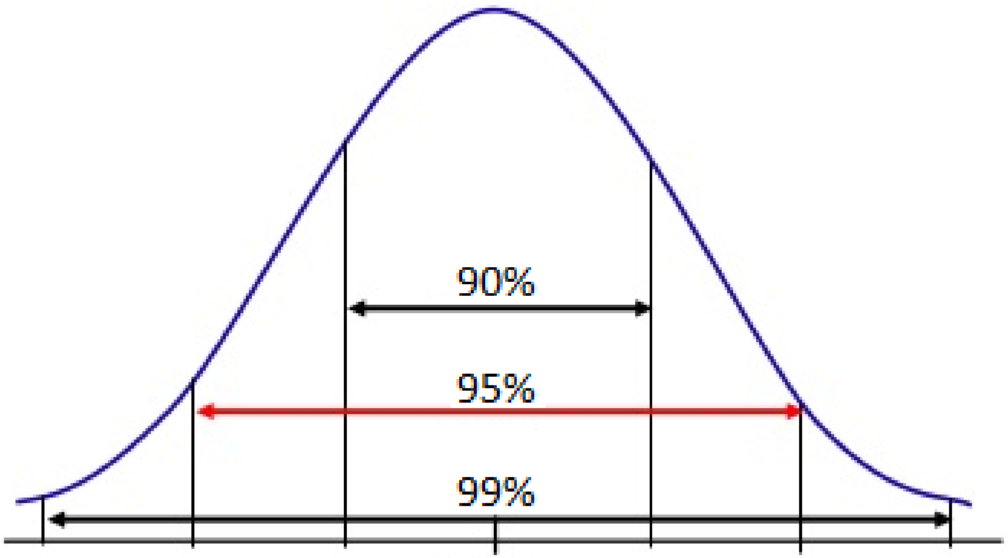
Un intervalo de confianza (IC) denota un rango dentro del cual puede encontrarse el parámetro, y el nivel de confianza que el intervalo contiene del parámetro.

Un intervalo de confianza tiene un límite inferior de confianza (LIC) y un límite superior de confianza (LSC).

Universidad de Medellín
Ciencia y Libertad

El coeficiente de confianza es el nivel de confianza que se tiene en el que el intervalo contenga el valor desconocido del parámetro.

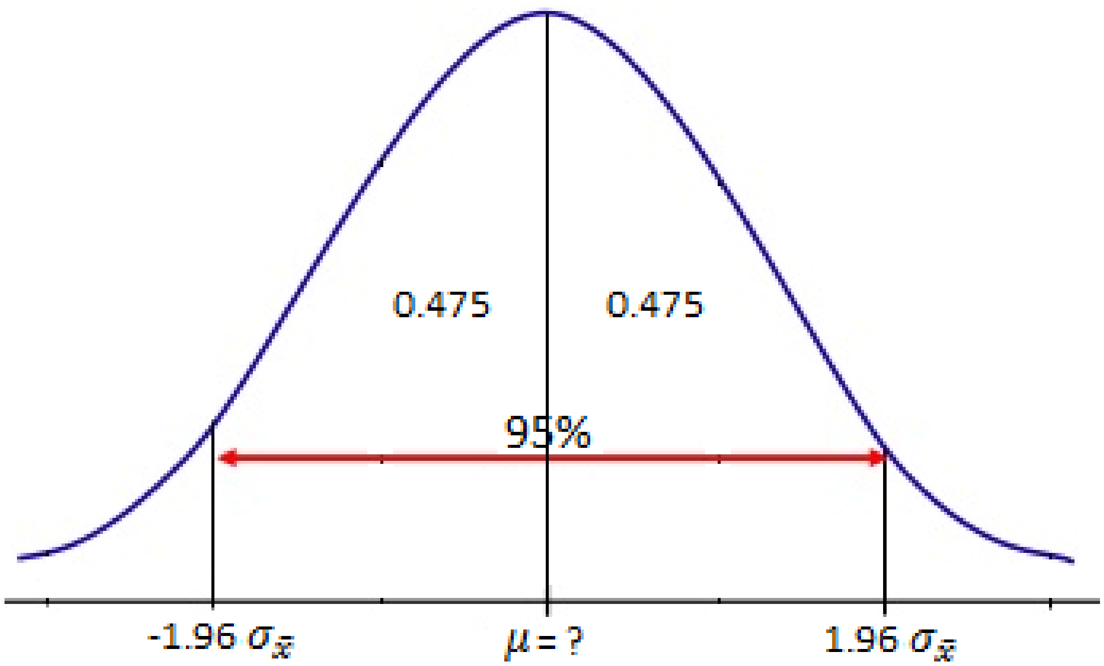
Los coeficientes de confianza más comúnmente utilizados son 99,95 y 90%.



Universidad de Medellín
Ciencia y Libertad

Para 95%

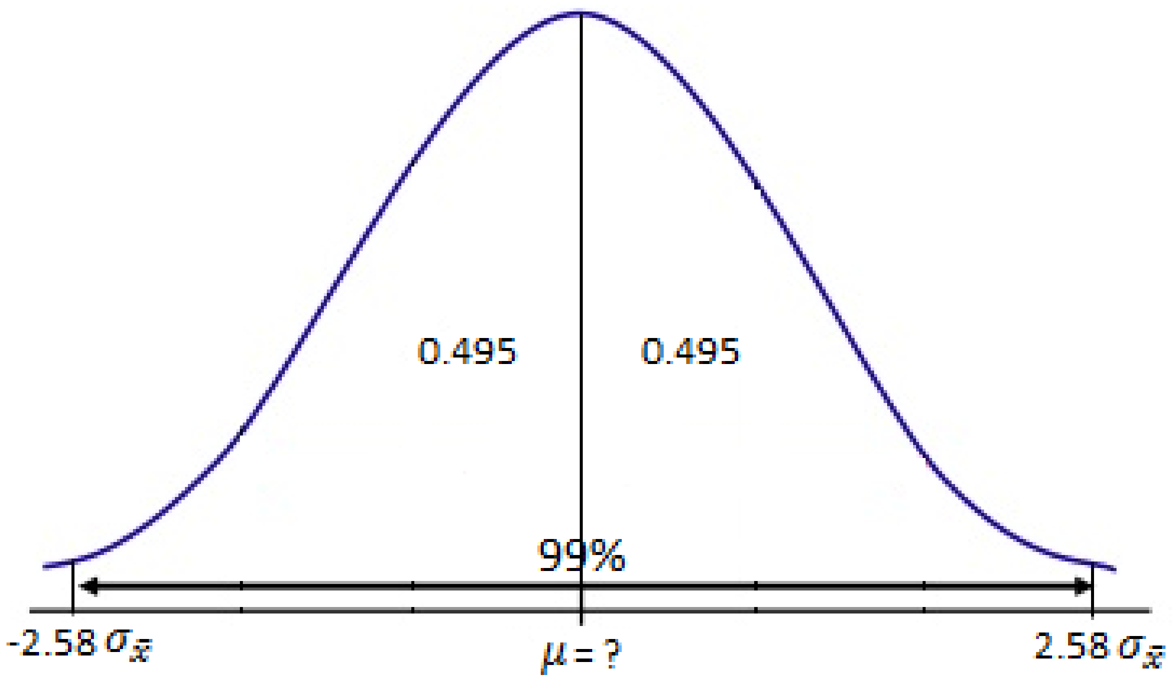
Debido a que la tabla Z contiene valores sólo para el área que está por encima o por debajo de la media, se debe dividir 95% por 2, produciendo 0.475. Luego se halla el valor de Z correspondiente a un área de 0.475 el cual es Z = 1.96. Así, para construir un intervalo de confianza del 95%, simplemente se especifica un intervalo de 1.96 errores estándar por encima y por debajo de la media muestral.



Universidad de Medellín
Ciencia y Libertad

Para 99%

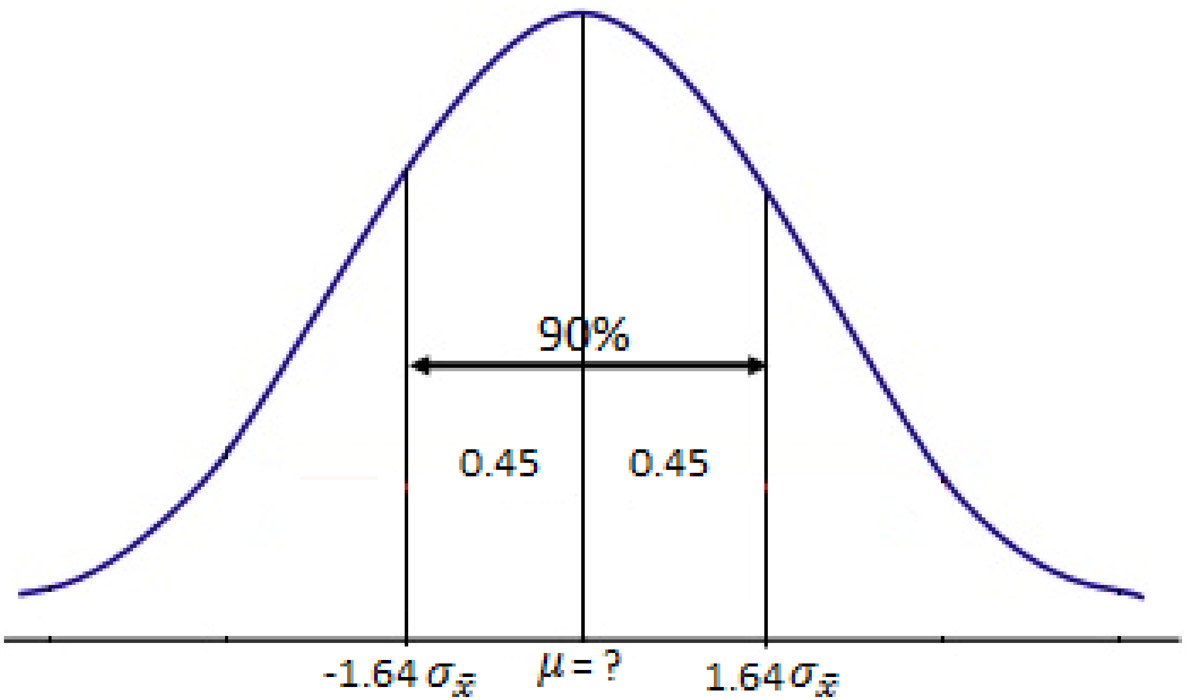
Debido a que la tabla Z contiene valores sólo para el área que está por encima o por debajo de la media, se debe dividir 99% por 2, produciendo 0.495. Luego se halla el valor de Z correspondiente a un área de 0.495 el cual es Z = 2.58. Así, para construir un intervalo de confianza del 99%, simplemente se especifica un intervalo de 2.58 errores estándar por encima y por debajo de la media muestral.



Universidad de Medellín
Ciencia y Libertad

Para 90%

Debido a que la tabla Z contiene valores sólo para el área que está por encima o por debajo de la media, se debe dividir 90% por 2, produciendo 0.45. Luego se halla el valor de Z correspondiente a un área de 0.45 el cual es Z = 1.64. Así, para construir un intervalo de confianza del 90%, simplemente se especifica un intervalo de 1.64 errores estándar por encima y por debajo de la media muestral.



Universidad de Medellín
Ciencia y Libertad

Intervalo de confianza para la media poblacional – muestras grandes $(n \geq 30)$

Universidad de Medellín
Ciencia y Libertad

Intervalo de confianza para estimar μ cuando σ es conocida

$$IC\ para\ estimar\ \mu = \bar{X} \pm Z\sigma_{\bar{x}}$$

Donde

$\bar{X}$ es la media muestral

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

n es el número de observaciones en la muestra

Z es el valor de Z para el coeficiente de confianza deseado

Cuanto debe sumarse o restarse, depende del nivel de confianza deseado.

Universidad de Medellín
Ciencia y Libertad

Un promotor inmobiliario desea construir un centro comercial. Él puede estimar en el área donde desea construir el ingreso promedio por familia como indicador de ventas esperadas. Una muestra de n = 100 familias da una media de $\bar{X}$ = 35500. Se asume que la desviación estándar poblacional es σ = 7200. Se asume un intervalo de confianza del 95%

$$IC\ para\ estimar\ \mu = \bar{X} \pm Z\sigma_{\bar{x}}$$

$$IC\ para\ estimar\ \mu = 35500 \pm (1.96)(\frac{7200}{\sqrt{100}})$$

$$\bar{X} - Z\sigma_{\bar{x}} \leq \mu \leq \bar{X} + Z\sigma_{\bar{x}}$$

$$34088.80 \leq \mu \leq 36911.20$$

Universidad de Medellín
Ciencia y Libertad

Interpretación:

1. El promotor tiene un 95% de confianza en que la media poblacional real desconocida esté entre $34088.80 \leq \mu \leq 36911.20$ aunque el valor real sigue siendo desconocido.
2. Si se construyen todos los ${}_{N}C_{n}$ intervalos de confianza, el 95% de ellos contendrá la media poblacional desconocida.

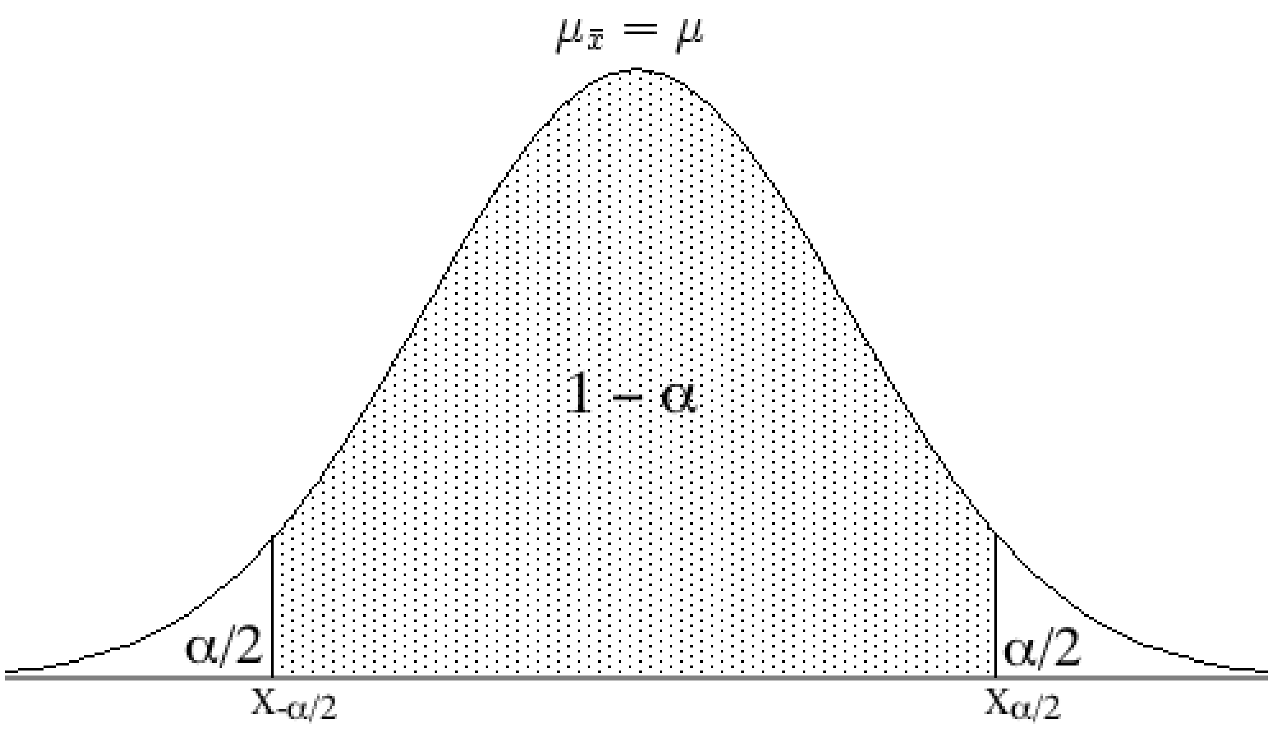
$$IC\ para\ estimar\ \mu = 35600 \pm (1.96)\left(\frac{7200}{\sqrt{100}}\right)$$

$$34188.80 \leq \mu \leq 37011.20$$

Esto por supuesto significa que el 5% de todos los intervalos estaría errado (no contendrían la media poblacional).

Universidad de Medellín
Ciencia y Libertad

Este 5% hallado como (1 – coeficiente de confianza), es denominado el valor alfa y representa la probabilidad de error. El valor alfa es la probabilidad de que cualquier intervalo dado no contenga la media poblacional.



Universidad de Medellín
Ciencia y Libertad

Intervalo de confianza para estimar μ cuando σ es desconocida

$$IC\ para\ estimar\ \mu = \bar{X} \pm Z S_{\bar{x}}$$

Donde

$\bar{X}$ es la media muestral

$S_{\bar{x}} = \frac{S}{\sqrt{n}}$

n es el número de observaciones en la muestra

Z es el valor de Z para el coeficiente de confianza deseado

Los intervalos con coeficientes de confianza pequeños son más precisos. Por lo tanto, si se toma el coeficiente de 95% en vez del de 99% se gana en precisión pero se pierde algo de confianza.

Universidad de Medellín
Ciencia y Libertad

Gerry Gerber, CPA, acaba de registrar las declaraciones de impuestos de sus clientes. Desea estimar la cantidad promedio que deben al servicio de renta interna. De los 50 clientes que seleccionó en su muestra, la cantidad promedio que se adeudaba era de 652.68. La desviación estándar de todos los clientes σ es desconocida, la desviación estándar de la muestra es s = 217.43. Se desea un nivel de confianza del 99%.

$$IC\ para\ estimar\ \mu = \bar{X} \pm Zs_{\bar{x}}$$

$$IC\ para\ estimar\ \mu = 652.68 \pm (2.58)(\frac{217.43}{\sqrt{50}})$$

$$\bar{X} - Zs_{\bar{x}} \leq \mu \leq \bar{X} + Zs_{\bar{x}}$$

$$573.35 \leq \mu \leq 732.01$$

Qué pasaría a este intervalo si el señor Gerber estuviera dispuesto a aceptar un nivel de confianza del 95%?

Universidad de Medellín
Ciencia y Libertad

$$IC\ para\ estimar\ \mu = 652.68 \pm (1.96)(\frac{217.43}{\sqrt{50}})$$

$$592.41 \leq \mu \leq 712.96$$

El nivel de confianza del 95% es más estrecho y ofrece mayor precisión. Sin embargo, el señor Gerber tuvo que abandonar algo de confianza para ganar en precisión.

Universidad de Medellín
Ciencia y Libertad

Ejercicio 7.1. Checkered Cabs planea comprar una flota de nuevos taxis para sus operaciones en Miami. La decisión depende de si el rendimiento del auto en consideración es por lo menos 27.5 millas por galón de gasolina. Los 36 carros que prueba la compañía reportan una media de 25.6 millas por galón, con una desviación estándar de 3.5 millas por galón. A un nivel de confianza del 99%, ¿Qué aconsejaría a Checkered que hiciera?

Ejercicio 7.2. Cien latas de 16 onzas de la salsa de tomate Jake's Mom's tiene un promedio de 15.2 onzas. La desviación estándar de la población en peso es de 0.96 onzas. ¿A un nivel de confianza de 95% las latas parecen estar llenas con un promedio de 16 onzas?

Ejercicio 7.3. Un estudio realizado por los profesores de una universidad de Kansas está diseñado para ofrecer inferencias sobre las tasas de desempleo por condado en Estados Unidos. Una muestra de 200 condados reporta una tasa promedio del 6.2%, con una desviación estándar de 1.7%. A un nivel de confianza del 90%, ¿cuál es el estimado de la tasa de desempleo promedio por condado en la nación?

Universidad de Medellín
Ciencia y Libertad

Intervalo de confianza para la media poblacional en el caso de muestras pequeñas ($n < 30$) – la distribución t

Cuando se utiliza una muestra pequeña, puede ser necesaria una distribución alternativa, la distribución t student. Específicamente la distribución t se utiliza cuando se cumplen estas tres condiciones:

1. La muestra es pequeña.

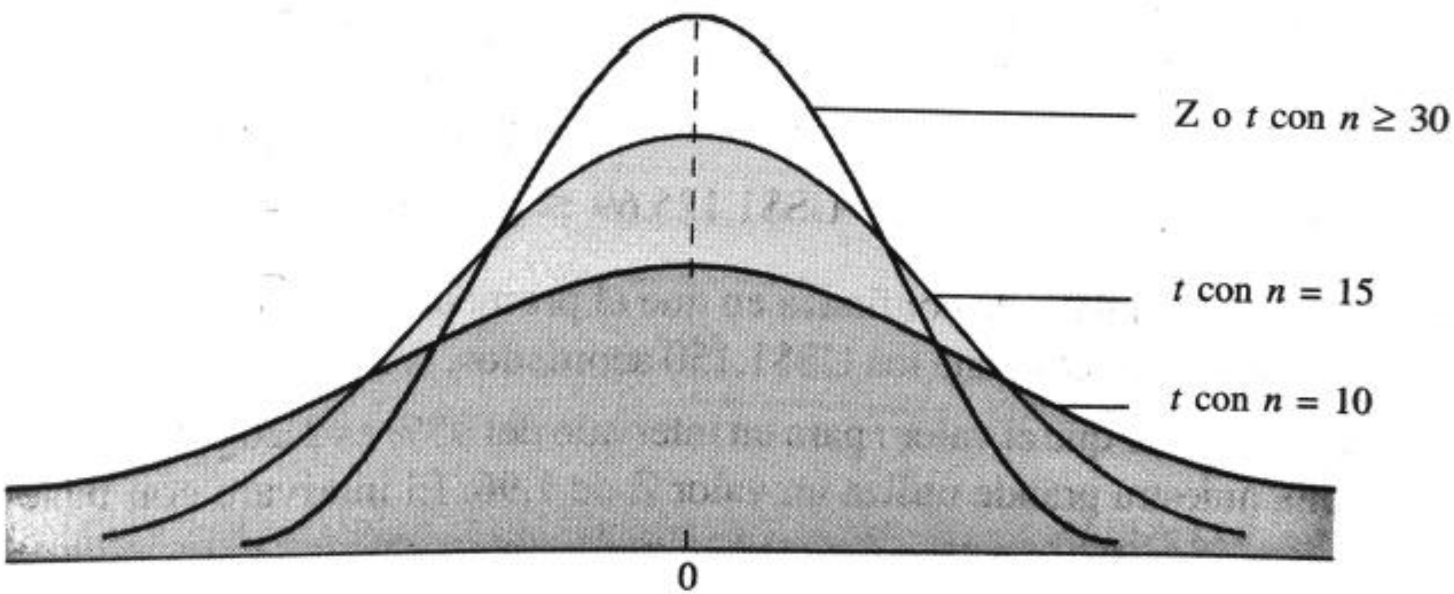
2. σ es desconocida

3. La población es normal o casi normal.

Si σ es conocida, la distribución Z se usa inclusive si la muestra es pequeña.

Universidad de Medellín
Ciencia y Libertad

Al igual que la distribución Z, la distribución t tiene una media de cero, es simétrica con respecto a la media y oscila entre $-\infty\ y + \infty$. Sin embargo, mientras que la distribución Z tiene una varianza de $\sigma^2 = 1$, la varianza de la distribución t es mayor que 1. Por lo tanto, es más plana y más dispersa que la distribución Z.



Varianza de la distribución t

$$\sigma^2 = \frac{n-1}{n-3}$$

Universidad de Medellín
Ciencia y Libertad

Grados de libertad: es el número de observaciones menos el número de restricciones impuestas sobre tales observaciones. Para nuestro caso será:

$$g.l. = n - 1$$

La distribución t producirá un intervalo más amplio que la distribución Z. Este ancho adicional es necesario debido a que se pierde algo de precisión porque σ es desconocida y debe estimarse.

$$t = \frac{\bar{X} - \mu}{S_{\bar{x}}}$$

El valor de t se encuentra en tablas, a partir del intervalo de confianza que se desea (ejemplo 95% es 0.95), y el número de grados de libertad (si n = 20, g.l. = 19), se toman en cuenta los intervalos de confianza para las pruebas de dos colas.

Universidad de Medellín
Ciencia y Libertad

Intervalo de confianza para estimar μ con distribución t

$$IC\ para\ estimar\ \mu = \bar{X} \pm t\frac{S}{\sqrt{n}}$$

Donde

$\bar{X}$ es la media muestral

S es la desviación estándar de la muestra

n es el número de observaciones en la muestra

t es el valor de t para el coeficiente de confianza deseado

Universidad de Medellín
Ciencia y Libertad

Una empresa de construcción fue culpada de inflar los comprobantes que registra para los contratos de construcción con el gobierno federal. El contrato estableció que un cierto tipo de trabajo debería promediar 1150. Por motivos de tiempo sólo 12 agencias del gobierno fueron llamados a dar testimonio ante la corte respecto a los comprobantes de la empresa. Si se descubrió a partir del testimonio que una media de 1275 y una desviación estándar de 235, ¿Un intervalo de confianza del 95% apoyaría el caso legal de la empresa? Se asume que los montos de los comprobantes son normales.

$$g.l. = n - 1 = 12 - 1 = 11$$

$$IC\ para\ estimar\ \mu = \bar{X} \pm t\frac{S}{\sqrt{n}}$$

$$IC\ para\ estimar\ \mu = 1275 \pm 2.201\frac{235}{\sqrt{12}}$$

$$1125.69 \leq \mu \leq 1424.31$$

La corte puede tener un 95% de confianza en que el promedio de todos los comprobantes está entre 1125.69 y 1424.31. Este intervalo contiene los 1150 acordados, fortaleciendo la defensa de la empresa.

Universidad de Medellín
Ciencia y Libertad

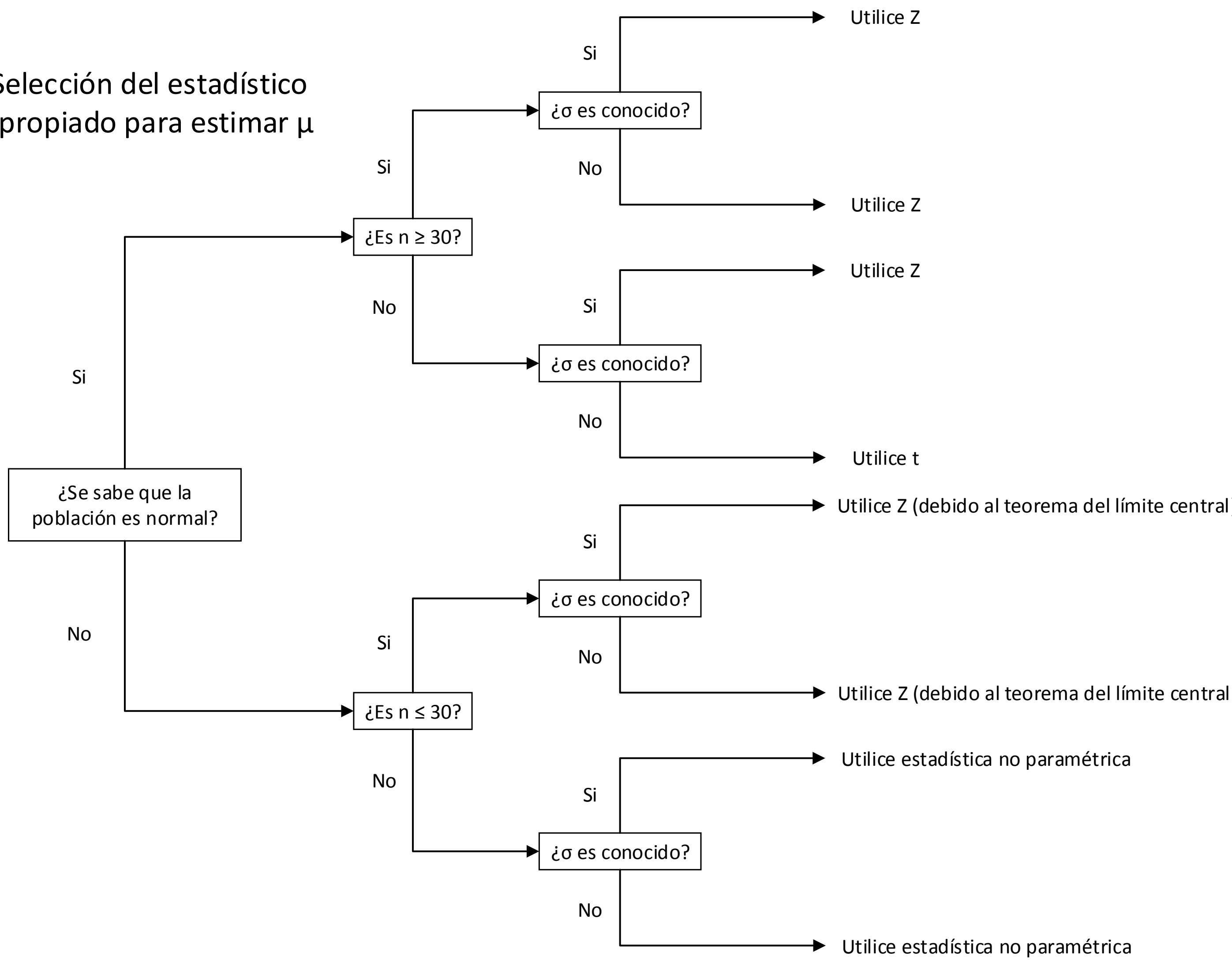
Ejercicio 7.4. El contrato laboral realizado entre United Auto Workers (UAW) y Ford Motor Company (FMC) requirió que la producción promedio para una sección de producción se realizara a 112 unidades por mes, por empleado. Surgieron desacuerdos entre ambas empresas respecto a si se mantenía este estándar o no. El contrato laboral especificó que si los niveles de producción promedio se reducían por debajo de la cantidad estipulada $\mu = 112$, a FMC se le permitiría tomar "acciones correctivas". Debido al costo implicado, sólo se evaluaron 20 trabajadores, resultando una media de 102 unidades. Se asume que se encontró una desviación estándar de 8.5 unidades y que los niveles de producción están distribuidos normalmente. ¿Un intervalo de confianza del 90% tiende a sugerir una contravención del contrato laboral, permitiendo así una acción correctiva?

Universidad de Medellín
Ciencia y Libertad

Ejercicio 7.5. The Lucky Lady, es una tertulia estudiantil popular, vende vasos de cerveza de 16 onzas. Diez estudiantes compran un total de 22 vasos, y utilizando su propia tasa de medida, estiman los contenidos promedio. La media muestral es de 15.2 onzas, con s = 0.86. ¿Con un nivel de confianza del 95% los estudiantes creen que su dinero lo vale? Interprete el intervalo.

Ejercicio 7.6. Las bonificaciones para 10 nuevos jugadores de la liga nacional de futbol se utiliza para estimar la bonificación promedio para todos los nuevos jugadores. La media muestral es de 65890 con s = 12300. ¿Cuál es su estimación para un intervalo del 90% para la media poblacional?

Selección del estadístico apropiado para estimar μ



Intervalo de confianza para la proporción poblacional

Universidad de Medellín
Ciencia y Libertad

Cuando las decisiones dependen de parámetros que son binarios, es decir, con sólo dos posibles categorías dentro de las cuales pueden clasificarse las respuestas, el parámetro de interés es la proporción poblacional.

Si la proporción poblacional π no es conocida, la proporción muestral p se utiliza como estimador de π.

Estimación del error estándar de la distribución de las proporciones muestrales

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

Donde
p es la proporción de éxitos en la muestra
n es el tamaño de la muestra

Universidad de Medellín
Ciencia y Libertad

Intervalo de confianza para estimar la proporción poblacional

$$IC\ para\ estimar\ \pi = p \pm ZS_p$$

Donde

p es la proporción de éxitos en la muestra

S_p es el error estándar de la distribución de las proporciones muestrales

Z es el valor de Z para el coeficiente de confianza deseado

Universidad de Medellín
Ciencia y Libertad

El gerente de una estación de televisión debe determinar en la ciudad que porcentaje de casas tiene más de un televisor. Una muestra aleatoria de 500 casas revela que 275 tienen dos o más televisores. ¿Cuál es el intervalo de confianza del 90% para estimar la proporción de todas las casas que tienen dos o más televisores?

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

$$S_p = \sqrt{\frac{(275/500)(1-\left(\frac{275}{500}\right))}{500}} = 0.022$$

$$IC\ para\ estimar\ \pi = p \pm ZS_p$$

$$IC\ para\ estimar\ \pi = 0.55 \pm (1.64)(0.022)$$

$$0.514 \leq \pi \leq 0.586$$

El gerente puede tener un 90% de confianza que entre el 51.4% y el 58.6% de las casas de la ciudad tienen más de un televisor.

Universidad de Medellín
Ciencia y Libertad

Ejercicio 7.7. Las empresas de búsqueda de ejecutivos se especializan en ayudar a las empresas a ubicar y asegurar talento para la alta gerencia. Tales empresas denominadas “cazadoras de cabezas” son responsables de la ubicación de muchos de los mejores directores ejecutivos de la nación. Business Week reportó recientemente que “uno de cada cuatro directores ejecutivos es una persona de fuera – un ejecutivo con menos de 5 años en la compañía que maneja”. Si en una muestra de 350 compañías de los Estados Unidos, 77 tienen directores ejecutivos de fuera, ¿Un intervalo del 99% de confianza apoyaría la afirmación?

Ejercicio 7.8. CNN informó que el 68% de todos los estudiantes de secundaria tenía computadores en sus casas. Si una muestra de 1020 estudiantes revela que 673 tienen computadores caseros, ¿Un intervalo del 99% apoya a CNN?

Ejercicio 7.9. La Asociación Nacional de Viajes tomó muestras de las personas que tomaban vacaciones en Irlanda para estimar la frecuencia con la cual los norteamericanos visitaban Emerald Isle. ¿Cuál es el intervalo de confianza del 96% para la proporción de turistas que son norteamericanos, si 1098 de los 3769 encuestados portaban pasaportes de Estados Unidos?

Control del ancho de un intervalo

Es preferible un intervalo más estrecho debido a la precisión adicional que proporciona. Hay dos métodos principales para lograr un intervalo más preciso:

1. Reducción del nivel de confianza (se sube la probabilidad del error).

   95% → probabilidad de error 5%

   99% → probabilidad de error 1%

2. Incremento del tamaño muestral: con este se puede reducir el error estándar $\frac{\sigma}{\sqrt{n}}$.

Universidad de Medellín
Ciencia y Libertad

Determinación del tamaño apropiado de la muestra

El tamaño de la muestra juega un papel importante al determinar la probabilidad de error así como en la precisión de la estimación. Una vez se ha seleccionado el nivel de confianza, dos factores importantes influyen en el tamaño muestral:

1. La varianza de la población σ^2 (está más allá del control del investigador).

2. El tamaño del error tolerable que el investigador está dispuesto a aceptar (depende de la precisión que se requiera).

El tamaño del error que un investigador puede tolerar depende de que tan crítico es el trabajo.

Tamaño de la muestra para estimar μ

$$n = \frac{Z^2\sigma^2}{(\bar{X} - \mu)^2}$$

Donde

Z nivel de confianza requerido

σ^2 es la varianza poblacional

$\bar{X}$ es la media muestral

μ es la media poblacional

$\bar{X} - \mu$ es el error tolerable

En el evento probable que σ^2 sea desconocido, puede estimarse mediante la desviación estándar muestral S, utilizando una muestra piloto de cualquier tamaño razonable $n \geq 30$.

$$n = \frac{Z^2(S)^2}{(error)^2}$$

Universidad de Medellín
Ciencia y Libertad

Se asume que un fabricante de reproductores de discos compactos desea construir un intervalo del 95% para el tamaño promedio de la pieza. Una muestra piloto a revelado una desviación estándar de 6mm. Si el error tolerable es de 2mm. ¿Qué tan grande debería ser la muestra?

$$n = \frac{Z^2(S)^2}{(error)^2}$$

$$n = \frac{(1.96)^2(6)^2}{(2)^2} = 34.5$$

El fabricante debería seleccionar una muestra de 35 piezas. De esta manera, un intervalo del 95% podría construirse para el tamaño promedio. El intervalo tendría un error no superior a 2mm.

Ejercicio 7.10. El propietario de un centro de esquí en el sur de Winsconsin está considerando comprar una máquina para hacer nieve y ayudarle a la madre naturaleza a proporcionar una base apropiada para los entusiastas esquiadores. Si el promedio de nevadas parece insuficiente, piensa que la máquina debería pagarse muy pronto por sí misma. Planea estimar las pulgadas promedio de nieve que cae en el área, pero no tiene idea de que tan grande debería ser la muestra. Sólo sabe que desea un 99% de confianza en sus hallazgos, que la desviación estándar es 3.5 y que el error no debe exceder 1 pulgada.

Universidad de Medellín
Ciencia y Libertad

Tamaño de la muestra para estimar π

Tamaño muestral para intervalos para la proporción poblacional

$$n = \frac{Z^2(\pi)(1-\pi)}{(p-\pi)^2}$$

Donde

Z nivel de confianza requerido

π es la proporción poblacional

p es la proporción muestral

$p - \pi$ es el error tolerable

En la fórmula se requiere el valor de π. Sin embargo, π es el parámetro que se desea estimar y es desconocido. Este problema puede tratarse de dos maneras:

1. Se podría tomar una muestra piloto para obtener un valor preliminar de π.
2. Se puede determinar que $\pi = 0.5$, para efectos de determinar el tamaño muestral.

Universidad de Medellín
Ciencia y Libertad

Wally Simpleton está postulado para gobernador. Él desea estimar dentro de 1 punto porcentual la proporción de personas que votarán por él. También desea tener el 95% de confianza de sus hallazgos. ¿Qué tan grande debería ser el tamaño muestral?

$$n = \frac{Z^2(\pi)(1-\pi)}{(p-\pi)^2}$$

$$n = \frac{(1.96)^2(0.5)(0.5)}{(0.01)^2} = 9604$$

Una muestra de 9604 votantes permitirá a Wally estimar π con un error de 1% y un nivel de confianza del 95%.

Ejercicio 7.11. El consejo de la ciudad está planeando una ley que prohíbe fumar en edificios públicos incluyendo restaurantes, tabernas y teatros. Sólo estará exenta la vivienda privada. Sin embargo, antes que dicha ley se lleve ante el consejo, este organismo desea estimar la proporción de residentes quienes apoyan dicho plan. La carencia de toda habilidad estadística obliga al consejo a contratarlo como consultor. Su primer paso será determinar el tamaño muestral necesario. Se le dice que su error no debe exceder el 2% y usted debe estar 95% seguro de sus resultados.

Ejercicio 7.12. Days Inn desea desarrollar un intervalo de confianza del 99% para estimar el número promedio de habitaciones ocupadas cada noche en sus localidades de toda la nación. ¿Cuántas noches deben incluirse en la muestra si se puede tolerar un error de 50 habitaciones y una muestra piloto revela que $s = 165$ habitaciones?

Ejercicio 7.13. Como empleado recién contratado en la división de mercadeo para un importante asunto sobre ventas minoristas, a usted se le ha asignado la tarea de estimar la proporción de consumidores que prefieren su producto al de la competencia. ¿Cuántos consumidores se deben tomar en la muestra si usted desea restringir el error al 10%, pero sin embargo desea proporcionar un nivel de confianza del 99%?

Propiedades de un buen estimador

Un estimador es el proceso mediante el cual se obtiene la estimación.

Una estimación es el resultado numérico del estimador.

Para desempeñarse de manera confiable los estimadores deben ser:

1. Estimador insesgado: un estimador es insesgado si la media de su distribución muestral es igual al parámetro correspondiente.

2. Estimador eficiente: dado un estimador insesgado, el estimador más eficiente es aquel que tenga la varianza más pequeña.

3. Estimador consistente: un estimador es consistente si, a medida que n aumenta, el valor estadístico se aproxima al parámetro.

4. Estimador suficiente: un estimador es suficiente si ningún otro estimador puede proporcionar más información sobre el parámetro.

Universidad de Medellín
Ciencia y Libertad