

Distribuciones muestrales

Estadística inferencial: involucra el uso de un estadístico para sacar una conclusión o inferencia sobre el parámetro (de la población) correspondiente.

El estadístico se utiliza como estimador del parámetro, esto depende de la muestra que se haya tomado.

Error de muestreo: es la diferencia entre el parámetro poblacional y el estadístico de la muestra utilizado para estimar el parámetro.

Aunque nunca se puede calcular realmente el tamaño del error de muestreo debido a que el parámetro de la población es desconocido, se debe ser consiente de que es probable de que exista un error de muestreo.

Se tienen los ingresos de 4 estudiantes universitarios, estos ingresos son de 100, 200 ,300 y 400 dólares.

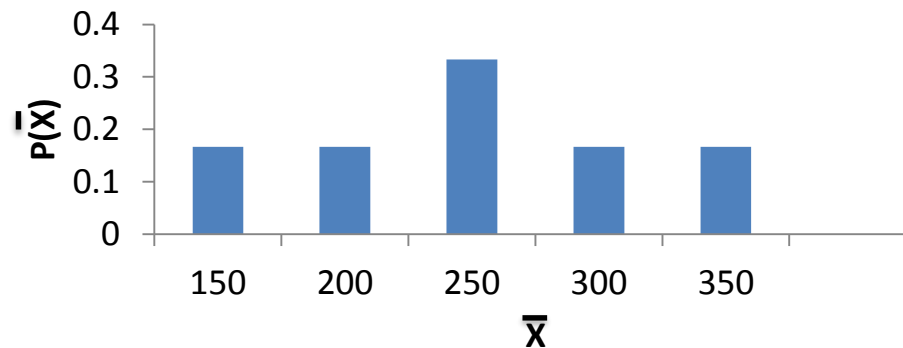
$$\mu = \frac{100 + 200 + 300 + 400}{4} = 250$$

Se puede seleccionar una muestra de $n = 2$ observaciones para estimar el μ “desconocido”, se podrá seleccionar aleatoriamente una muestra de ${}_4C_2 = 6$ posibles muestras.

Muestra	Elementos muestrales X_i	Medias muestrales \bar{X}
1	100, 200	150
2	100, 300	200
3	100, 400	250
4	200, 300	250
5	200, 400	300
6	300, 400	350

Distribución muestral: es una lista de todos los valores posibles para un estadístico y la probabilidad relacionada con cada valor.

Medias muestrales \bar{X}	# de muestras que dan \bar{X}	Probabilidad de $P(\bar{X})$
150	1	1/6
200	1	1/6
250	2	2/6
300	1	1/6
350	1	1/6
		1



Media de las medias muestrales

Una distribución muestral de las medias muestrales es simplemente una lista de todas las medias posibles. Éstas, al igual que cualquier lista de números tienen una media denominada la media de las medias muestrales o la gran media.

$$\bar{\bar{X}} = \frac{\sum \bar{X}}{K}$$

Donde

K es el número de muestras en la distribución muestral.

La media de la distribución muestral $\bar{\bar{X}}$ es igual a la media poblacional original μ .

No se debe confundir n que es el número de observaciones en una sola muestra, con K que es el número de muestras en la distribución normal.

La media de las muestras muestrales será:

$$\bar{X} = \frac{150 + 200 + 250 + 250 + 300 + 350}{6} = 250$$

$$\bar{X} = \mu = 250$$

La varianza y el error estándar de la distribución muestral de las medias muestrales

Varianza: mide la dispersión de las observaciones (medias muestrales) alrededor de su media (la gran media).

$$\sigma_{\bar{X}}^2 = \frac{\sum(\bar{X} - \bar{\bar{X}})^2}{K} = \frac{\sum(\bar{X} - \mu)^2}{K}$$

Donde

$\bar{\bar{X}}, \mu$ es la media de la distribución muestral

\bar{X} cada una de las observaciones, medias muestrales

K número de medias muestrales

$$\sigma_{\bar{X}} = \sqrt{\sigma_{\bar{X}}^2}$$

Debido a que la diferencia entre \bar{X} y μ es el error de muestreo, toda medida de la tendencia de la media muestral a desviarse de μ se le denomina acertadamente **error estándar**. Por lo tanto, el error estándar $\sigma_{\bar{X}}$ mide la tendencia a sufrir del error de muestreo en el esfuerzo por estimar μ .

La varianza y el error estándar de la distribución muestral de las medias muestrales serán:

$$\sigma_{\bar{X}}^2 = \frac{(150 - 250)^2 + (200 - 250)^2 + (250 - 250)^2 + (250 - 250)^2 + (300 - 250)^2 + (350 - 250)^2}{6}$$

$$\sigma_{\bar{X}}^2 = 4.167$$

$$\sigma_{\bar{X}} = \sqrt{4.167} = 64.55$$

Ejercicio 6.1. Las ventas en miles de dólares para East Coast Manufacturing (ECM) durante los últimos 5 meses fueron de 68, 73, 65, 80 y 72. Asumiendo que estos cinco meses constituyen la población, la media claramente es $\mu=71.6$. Como director de marketing de ECM, se desea estimar este μ “desconocido” tomando una muestra de tamaño $n = 3$. Se espera que el error de muestreo que es probable que ocurra sea relativamente pequeño. Realice la distribución muestral y haga comentarios sobre el posible error de muestreo.

Una aproximación cercana a la varianza y al error estándar se puede obtener así:

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \qquad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Esto solamente es apropiado si el muestreo se realiza con reemplazo o si la muestra se toma de una población muy grande y además se conoce la varianza poblacional σ^2 .

Si el muestreo se realiza sin reemplazo y si el tamaño de la muestra es más del 15% de la población, $n > 0.05N$, debe aplicarse el factor de corrección para poblaciones finitas (fpc).

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

Donde:

$\sqrt{\frac{N-n}{N-1}}$ es el fpc

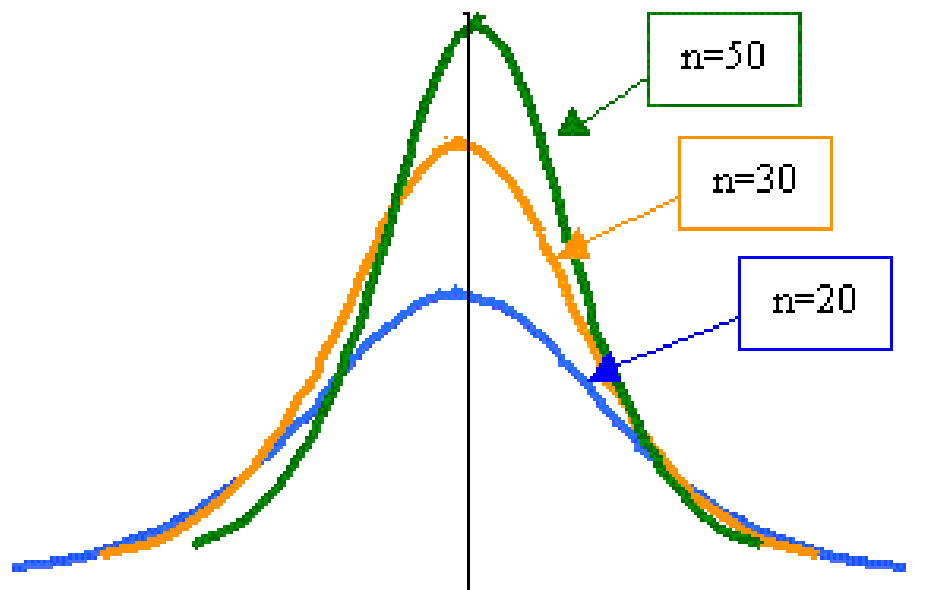
Si n es pequeño con respecto a N (menos del 5%), el fpc se aproxima a 1.

A medida que n aumenta, $\sigma_{\bar{X}}$ disminuye. Las muestras más grandes llevan a errores de muestreo más pequeños.

Teorema del límite central

A medida que n se vuelve más grande, la distribución de las medias muestrales se aproximará a una distribución normal con una media $\bar{X} = \mu$ y un error estándar $\sigma_{\bar{X}} = \sigma/\sqrt{n}$.

La regla general es que si n es por lo menos 30, el teorema del límite central asegurará una distribución normal en las medias muestrales incluso si la población no es normal.



Distribuciones de medias muestrales

Uso de la distribución normal

Muchas decisiones dependen de una muestra completa y no solo de una observación. En este caso, la fórmula de conversión debe alterarse para explicar el hecho en el cual se está interesado, no solo en una observación X sino en la media de varias observaciones \bar{X} . Por tanto, cuando se hace el muestreo, la fórmula de conversión se vuelve

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

Donde:

Z es la desviación normal

\bar{X} es la media de varias observaciones (n observaciones).

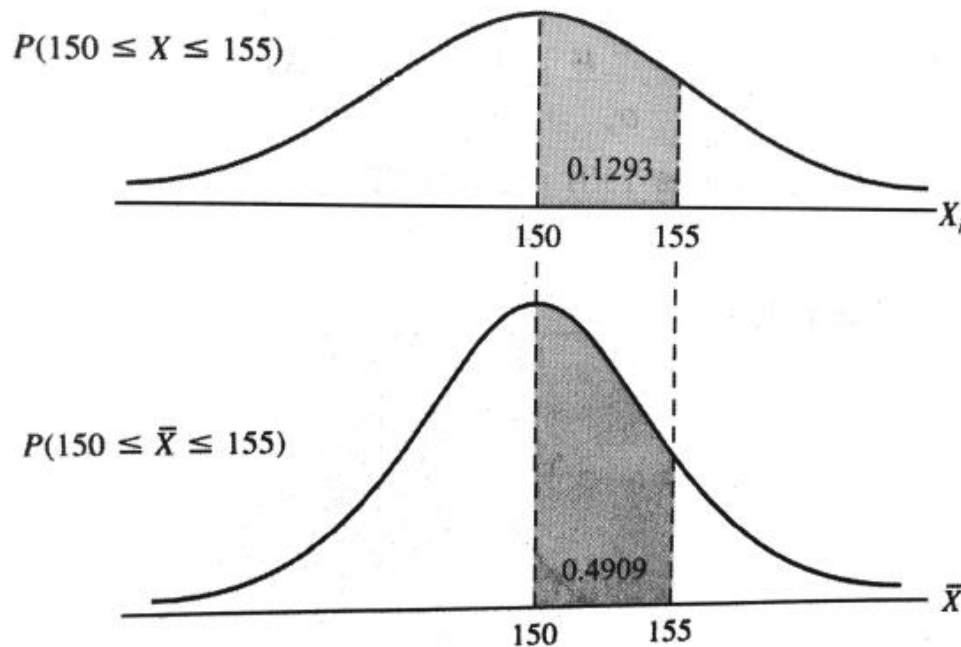
μ es la media

$\sigma_{\bar{X}}$ es el error estándar de la distribución muestral

El valor de interés en el numerador no es una observación única X_i , sino la media de n observaciones. Además, el denominador no es la desviación estándar poblacional σ , sino el error estándar de la distribución muestral $\sigma_{\bar{X}}$. Por ejemplo, en lugar de querer determinar la duración de una sola llamada, se puede calcular la probabilidad de que la media de n llamadas dure un cierto período de tiempo.

TelCom Satellite presta servicios de comunicación a los negocios del área metropolitana de Chicago. Los funcionarios de la compañía han aprendido que la transmisión de satélite promedio es de 150 segundos, con una desviación estándar de 15 segundos. Los tiempos parecen estar distribuidos normalmente.

TelCom desea ahora conocer la probabilidad de que la media de $n = 50$ llamadas esté entre 150 y 155 segundos.



$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{155 - 150}{\frac{15}{\sqrt{50}}} = 2.36$$

$$P(150 \leq \bar{X} \leq 155) = 0.4909$$

Ejercicio 6.2. TelCom Satellite presta servicios de comunicación a los negocios del área metropolitana de Chicago. Los funcionarios de la compañía han aprendido que la transmisión de satélite promedio es de 150 segundos, con una desviación estándar de 15 segundos. Los tiempos parecen estar distribuidos normalmente.

TelCom planea instalar nuevos equipos que mejorarían la eficiencia de sus operaciones. Sin embargo, antes que los ejecutivos puedan decidir si dicha inversión será eficaz en función de los costos, deben determinar la probabilidad de que la media de una muestra de $n = 35$:

1. Esté entre 145 y 150 segundos.
2. Sea mayor que 145 segundos.
3. Sea menor que 155 segundos.
4. Esté entre 145 y 155 segundos.
5. Sea mayor que 155 segundos.

Distribución de las proporciones muestrales

Cuando no se trabaja con muestras sino con la proporción, se quiere es determinar la proporción de éxitos y la proporción de fracasos. En estos casos se utiliza la proporción muestral p para estimar el parámetro desconocido π .

El proceso de las proporciones muestrales es muy similar al de las medias. De cualquier población es posible obtener muchas muestras diferentes de un tamaño dado. Cada muestra tendrá su propia proporción de “éxitos” p . Sin embargo, al igual que con las medias, el valor esperado de la distribución muestral de las proporciones muestrales será igual a la proporción de éxitos en la población: $E(p) = \pi$.

Valor esperado de la distribución muestral:

$$E(p) = \frac{\sum p}{K}$$

Donde

K es el número de proporciones en la distribución muestral.

p es la proporción de éxitos de cada muestra

Error estándar:

$$\sigma_p = \sqrt{\frac{(\pi)(1 - \pi)}{n}}$$

Donde

π es la proporción de éxitos en la población

n es el tamaño de la muestra

N es el tamaño de la población

De la misma manera que con las medias, si $n > 0.05N$, se requiere el fpc y el error estándar se vuelve

$$\sigma_p = \sqrt{\frac{(\pi)(1 - \pi)}{n}} \sqrt{\frac{N - n}{N - 1}}$$

Lugget Furniture pregunta a toda la población $N = 4$ clientes si vieron el anuncio publicitario de Lugget en el periódico de esta mañana. Se registró una respuesta de “sí” como éxito y de “no” como fracaso. Los cuatro clientes respondieron S_1, N_2, N_3 y S_4 . La proporción poblacional de éxitos es $\pi = 0.5$. Se tomaron muestras de tamaño $n = 2$, la proporción de éxitos es la siguientes:

X_i	Número de éxitos	p (proporción de éxitos)
S_1, N_2	1	0.5
S_1, N_3	1	0.5
S_1, S_4	2	1
N_2, N_3	0	0
N_2, S_4	1	0.5
N_3, S_4	1	0.5
		3

$$E(p) = \frac{\sum p}{K} = \frac{3}{6} = 0.5 = \pi$$

$$n > 0.05N = 2 > 0.05(4) = 2 > 0.2$$

$$\sigma_p = \sqrt{\frac{(\pi)(1 - \pi)}{n}} \sqrt{\frac{N - n}{N - 1}}$$

$$\sigma_p = \sqrt{\frac{(0.5)(1 - 0.5)}{2}} \sqrt{\frac{4 - 2}{4 - 1}} = 0.289$$

La desviación normal para las proporciones, permite determinar las probabilidades para las proporciones muestrales.

$$Z = \frac{p - \pi}{\sigma_p}$$

Donde

p es la proporción de éxitos de la muestra

π es la proporción de éxitos en la población

σ_p es el error estándar de la proporción

Ejercicio 6.3. Bellabs adquiere componentes para sus teléfonos celulares en lotes de 200 de una firma en Palo Alto. El componente tiene una tasa de defectos del 10%. Una política establecida recientemente por Bellabs establece que si el siguiente envío tiene:

1. Más del 12% de defectos, definitivamente buscará otro proveedor.
2. Entre el 10 y el 12% de defectos, considerará un nuevo proveedor.
3. Entre el 5 y el 10% de defectos, definitivamente no conseguirá un nuevo proveedor.
4. Menos del 5% de defectos, incrementará sus pedidos.

¿Cuál es la decisión más probable que tome Bellabs?

Métodos de muestreo

Una muestra que tergiverse la población presentará un error de muestreo y producirá estimados imprecisos de los parámetros de la población. Hay dos fuentes básicas de error de muestreo, la mala suerte y el sesgo muestral.



Muestreo aleatorio simple: tomar una muestra aleatoria simple garantiza que cada muestra de algún tamaño dado tenga la misma probabilidad de ser seleccionada. Se pueden generar sacando los elementos al azar (de un sombrero) o con números aleatorios.

Muestreo sistemático: se forma seleccionando cada i ésimo ítem de la población. La población debe ordenarse o enumerarse en forma aleatoria. El peligro principal que debe evitarse es la ocurrencia de un patrón en el ordenamiento de la población.





Muestreo estratificado: se toma una muestra estratificada forzando las proporciones de la muestra de cada estrato para que esté conforme al patrón poblacional. Se emplea comúnmente cuando la población es heterogénea, o disímil, aunque ciertos subgrupos homogéneos puedan aislarse.

Muestreo por conglomerados: consiste en dividir toda la población en conglomerados, o grupos, y luego seleccionar una muestra de estos conglomerados. Todas las observaciones de estos conglomerados seleccionados están incluidas en la muestra.

