Estimación de Parámetros

Jhon Jairo Padilla A., PhD.

Inferencia Estadística

- La inferencia estadística puede dividirse en dos áreas principales:
  - Estimación de Parámetros
  - Prueba de Hipótesis

Estimación de Parámetros

- Modelado de sistemas:
  - Se posee un **conjunto de muestras** de un experimento aleatorio
  - Se desea obtener un valor estimado de los parámetros del sistema (valores con respecto a la **población total**)
- A el procedimiento usado para obtener los parámetros de la población total se le llama **Estimación de Parámetros.**
- En este procedimiento se requiere determinar la cercanía de la estimación con la realidad. Para esto se utilizan los **Intervalos de Confianza**.

Estimación de Parámetros

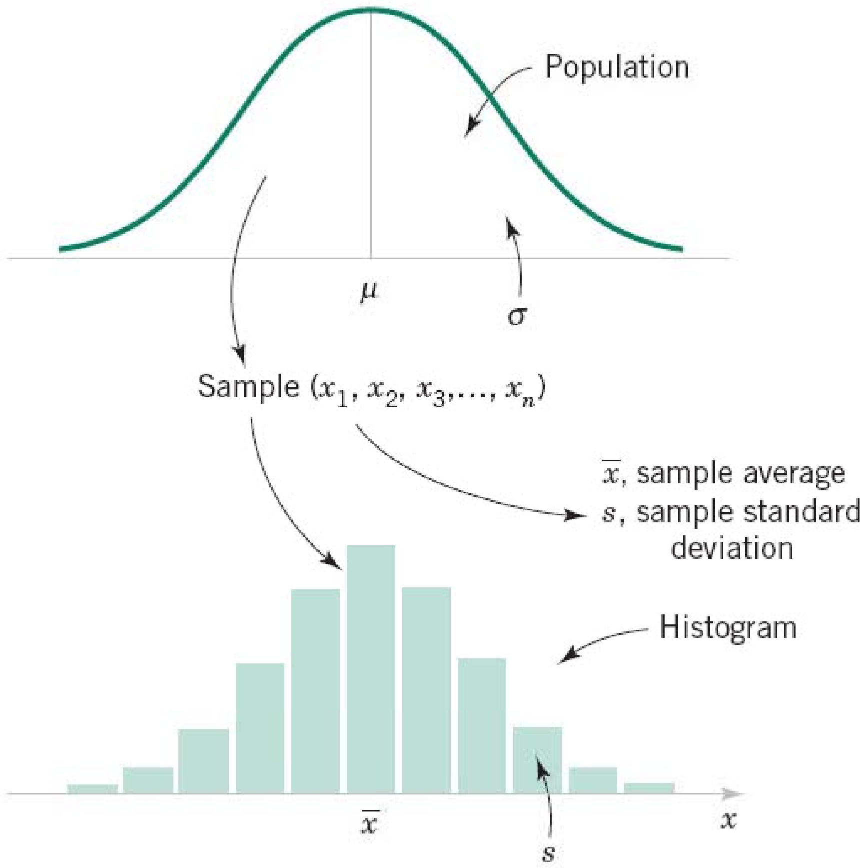


Figure 6-3 Relationship between a population and a sample.

Prueba de Hipótesis

- Se desea comparar dos tratamientos (métodos, procedimientos, mecanismos, funciones, etc) diferentes.
- Ejemplo:
  - Se tiene un proceso químico.
  - Un ingeniero puede usar dos temperaturas diferentes en el mismo proceso (t_1, t_2)
  - El ingeniero conjetura que t_1 produce rendimientos más altos que t_2.
  - El ingeniero asume una hipótesis a comprobar: *"El rendimiento medio utilizando la temperatura t_1 es mayor que el rendimiento medio utilizando la temperatura t_2"*
- No se hace énfasis en la estimación de los rendimientos; más bien, la atención se centra en sacar conclusiones acerca de una hipótesis propuesta.

ESTIMACION DE PARÁMETROS

Muestreo Aleatorio

- Se requiere tomar unas muestras de una población para obtener un modelo estadístico
- Recordemos:
  - **Población:** Totalidad de las observaciones que son motivo de interés
  - **Tamaño de la población:** Número de observaciones que hay en la población. Esta puede ser **finita y discreta** (Ej: Número de botellas con llenado incompleto en un día en una embotelladora) o **infinita y contínua** (Ej: Mediciones posibles del porcentaje de monóxido de carbono en un día en una calle).
  - A toda población se la puede modelar mediante una **distribución de probabilidad.**

Razones para Muestrear

- En la mayoría de ocasiones es imposible o poco práctico observar la población completa:
  - Podría requerirse gran cantidad de tiempo
  - Sería extremadamente costoso
  - Al momento de tomar una decisión podría no existir toda la población

Muestras

- Una **muestra** es un subconjunto de observaciones que se seleccionan de una población.
- Para que las inferencias sean **válidas**, la muestra debe ser **representativa** de la población.
- **Error común:** Tomar las muestras más sencillas de obtener. Como resultado habrá un error en el parámetro de interés (Hay **Sesgos** en la muestra).
- La toma de las muestras debe ser aleatoria.
- Cada observación de la muestra es el valor observado de una **variable aleatoria**.

Características del experimento

- Sea X una v.a. que representa el resultado de una selección de una observación de una población.
- Sea f(x) que denota la f.d.p. de X
- Supóngase que **cada observación** de la muestra se obtiene de forma **independiente**, bajo las mismas condiciones.
- Se hacen **n observaciones**. La v.a. X_i representa la observación en la repetición i. Se obtienen los valores numéricos $x_1, x_2, \ldots, x_n$
- Las observaciones realizadas tienen una misma distribución de probabilidad ya que fueron tomadas de forma independiente y bajo condiciones idénticas
- Por tanto, la función de distribución de probabilidad conjunta es

$$f_{X_1 X_2 \ldots X_n}(x_1, x_2, \ldots x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \ldots . f_{X_n}(x_n)$$

Muestra aleatoria

- Las variables aleatorias (X_1, X_2, …,X_n) son una muestra aleatoria de tamaño n si:
  - Las X_i son variables aleatorias independientes
  - Cada X_i tiene la misma distribución de probabilidad

Estadísticos

- Ejemplo:
  - Supóngase que se quiere establecer la proporción de la población de Colombia que prefiere una marca de refresco particular.
  - Sea p que representa el valor desconocido de esta proporción
  - Se selecciona una muestra aleatoria para hacer una inferencia respecto a p (no es práctico preguntar a cada individuo de la población).
  - Se obtiene una proporción observada p’
  - p’ se obtiene dividiendo el número de individuos que prefieren la marca de refresco entre el número total de la muestra (n).
  - p’ depende del número de valores observados (p’ varía de una muestra a otra)
  - Luego, p’ es una variable aleatoria y se conoce como **estadístico**.

Definición

- Un **estadístico** es cualquier función de las observaciones de una muestra aleatoria.

- Ejemplos:
  - Media muestral
  - Varianza muestral
  - Desviación estándar muestral

Estadísticos

- Un estadístico es una variable aleatoria
- Tiene una distribución de probabilidad, llamada **Distribución de muestreo**.
- Utilidad:
  - Se usan para obtener **estimaciones puntuales** de parámetros como: *media poblacional y varianza poblacional*.
- El parámetro de interés se representa por θ.
- El valor numérico de un estadístico muestral se usa como la **estimación puntual**.

Definición

- Sea X una v.a. con distribución de probabilidad f(x)
- Sea que f(x) está caracterizada por un parámetro desconocido θ.
- Sea X_1, X_2, …,X_n una muestra aleatoria de tamaño n.
- Al estadístico $\hat{\Theta} = h(X_1, X_2, ..., X_n)$ se le llama un **estimador puntual** de θ.
- $\hat{\Theta}$ es una v.a., ya que es función de v.a.'s.
- Al seleccionar la muestra, $\hat{\Theta}$ toma un valor numérico particular $\hat{\theta}$ llamado **estimación puntual** de θ.

Ejemplo

- Suponga una v.a. X que tiene una distribución normal con una media desconocida μ (media poblacional).
- La *media muestral* $\hat{\mu}$ es un **estimador puntual** de la media poblacional desconocida. Es decir, $\hat{\mu} = \bar{X}$.
- Si $x_1 = 25$, $x_2 = 30$, $x_3 = 29$, $x_4 = 31$ entonces la estimación puntual de μ es

$$\hat{\mu} = \bar{x} = \frac{25 + 30 + 29 + 31}{4} = 28.75$$

Parámetros comunes de estimación

Parámetro	**Estimación razonable**
Media de una población (μ)	Media muestral: $\hat{\mu} = \overline{X}$
Varianza de una población (σ^2) ó la desviación estándar (σ)	Varianza muestral: $\hat{\sigma}^2 = s^2$
Proporción *p* de elementos de una población que pertenecen a una clase de interés	Proporción muestral: $\hat{p} = \dfrac{x}{n}$ donde x es el número de elementos de una muestra de n elementos que pertenecen a la clase de interés
Diferencia de las medias de dos poblaciones: μ_1-μ_2	$\hat{\mu}_1 - \hat{\mu}_2 = \overline{x}_1 - \overline{x}_2$
La diferencia en las proporciones de dos poblaciones: p_1-p_2	$\hat{p}_1 - \hat{p}_2$

Objetivo de una estimación puntual

- Seleccionar, con base en los datos muestrales, un solo número que sea el valor más recomendable de θ.
- **Nota:** Puede haber varias opciones diferentes de estimadores de un parámetro.
- **Ejemplo:**
  - Parámetro a estimar: Media de una población
  - Posibles estimadores puntuales:
    - Media muestral
    - Mediana Muestral
    - Promedio de las observaciones menor y mayor de la muestra
  - Cuál será el mejor? Se requieren métodos para comparar estimadores.

Ejemplo

- Suponga que se toma una muestra aleatoria de tamaño n=10 de una población normal y se obtienen los datos de la tabla. Posibles estimadores son:
  - Media muestral $\bar{x} = \frac{\sum x_i}{n} = 11.04$
  - Mediana muestral $\tilde{x} = \frac{10.3 - 11.6}{2} = 10.95$
  - ¿Cuál es mejor?....

Valores de x
12.8
9.4
8.7
11.6
13.1
9.8
14.1
8.5
12.1
10.3

PROPIEDADES DE LOS ESTIMADORES

Estimador Insesgado

- El estimador puntual $\hat{\Theta}$ es un estimador insesgado del parámetro θ si

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

- Si el estimador no es insesgado, entonces, a la diferencia,

$$E(\hat{\Theta}) - \theta$$

se le llama el **sesgo** del estimador $\hat{\Theta}$.

Varianza de un estimador puntual

- Los dos estimadores son insesgados (tienen su centro en el valor real del parámetro estimado)
- El estimador que tenga menor varianza tendrá mayores posibilidades de estar cerca del valor estimado.

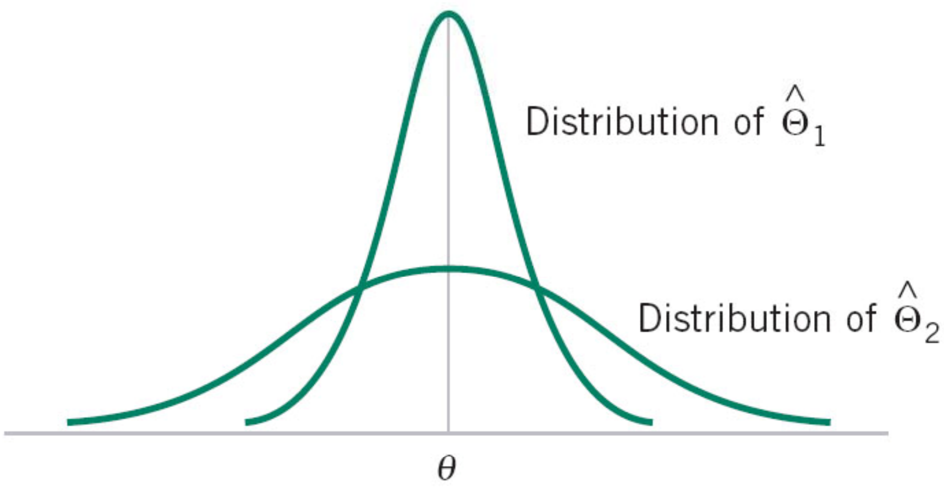


Figure 7-5 The sampling distributions of two unbiased estimators $\hat{\Theta}_1$ and $\hat{\Theta}_2$.

Escogencia de un estimador

- Si se consideran todos los estimadores insesgados de θ, al que tiene la varianza menor se le llama **estimador insesgado de varianza mínima** (MVUE).
- El MVUE es el estimador que tiene mayores posibilidades de estar cerca de θ.
- Si se desconoce el MVUE, podría usarse el principio de varianza mínima para elegir entre los posibles estimadores.

Caso importante: Distribución Normal

- Si X_1, X_2, X_3,…,X_n es una muestra aleatoria de tamaño ***n*** de una distribución normal con media μ y varianza σ^2, entonces la media muestral es el MVUE de μ.

Error estándar

- Da una idea de la precisión de la estimación.
- El error estándar de un estimador $\hat{\Theta}$ es su desviación estándar, dada por $\sigma_{\hat{\Theta}} = \sqrt{V(\hat{\Theta})}$.
- Si el error estándar incluye parámetros desconocidos que pueden estimarse, entonces la sustitución de dichos valores en $\sigma_{\hat{\Theta}}$ produce un error estándar estimado, denotado por $\hat{\sigma}_{\hat{\Theta}}$.

Caso: Distribución Normal

- Suponga que se hace un muestreo de una distribución normal con media μ y varianza σ^2. Entonces la distribución de $\overline{X}$ es normal con media μ y varianza σ^2/n, por lo que el error estándar de $\overline{X}$ es

$$\hat{\sigma}_{\bar{X}} = \frac{s}{\sqrt{n}}$$

Desviación estándar muestral

- Además, se puede suponer razonablemente que el valor real del parámetro está entre dos errores estándar de la estimación

Ejemplo

Medidas de conductividad térmica
41.60
41.48
42.34
41.95
41.86
42.18
41.72
42.26
41.81
42.04

- Un artículo del *Journal of Heat Transfer* describía un nuevo método para medir la conductividad térmica del hierro Armco. Utilizando una temperatura de 100°F y una alimentación de energía de 550W, se obtuvieron las 10 mediciones de la conductividad térmica de la tabla.
- Estimación puntual: $\overline{x} = \frac{\sum x_i}{n} = 41,924$
- Error estándar:

$$\hat{\sigma}_{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{0,284}{\sqrt{10}} = 0,0898$$

El valor medio real estará en el Intervalo 41,924$\pm$0,1796 (media mas/menos dos veces el error estándar)

El error estándar es el 0,2% de la media

Error cuadrado medio de un estimador

- Cuando se utilizan estimadores sesgados, es importante el error cuadrado medio del estimador.
- El error cuadrado medio de un estimador $\hat{\Theta}$ del parámetro θ se define como

$$MSE(\hat{\Theta}) = E(\hat{\Theta} - \theta)^2$$

$$MSE(\hat{\Theta}) = V(\hat{\Theta}) + sesgo^2$$

Criterio de comparación

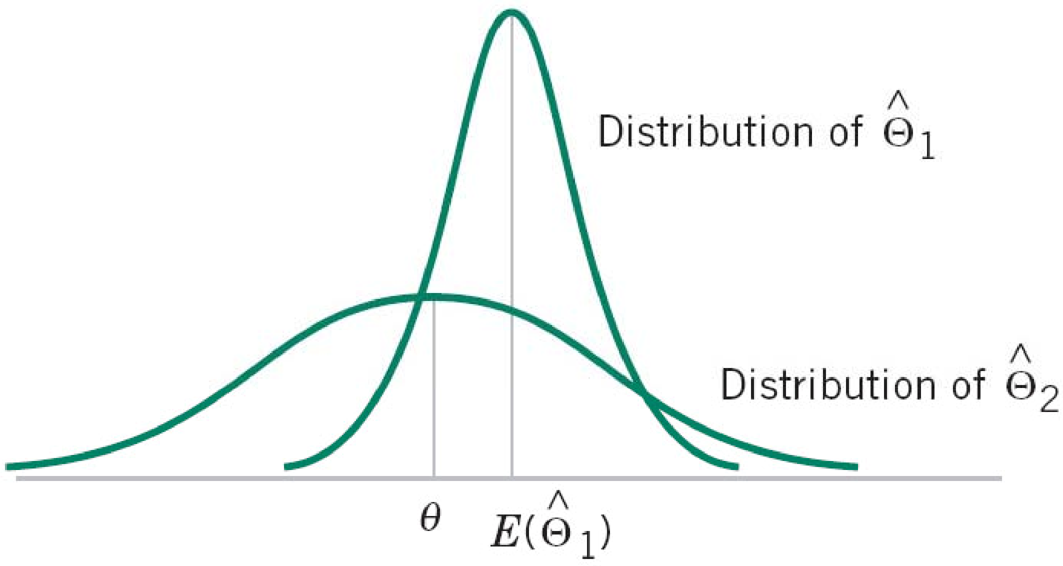
- El error cuadrado medio es un criterio de comparación de dos estimadores.
- Sean $\hat{\Theta}_1$ y $\hat{\Theta}_2$ dos estimadores del parámetro θ y sean MSE($\hat{\Theta}_1$) y MSE($\hat{\Theta}_2$) los errores cuadrados medios de θ_1 y θ_2. Entonces la eficiencia relativa de $\hat{\Theta}_2$ respecto a $\hat{\Theta}_1$ se define como

$$\frac{MSE(\hat{\Theta}_1)}{MSE(\hat{\Theta}_2)}$$

- Si esta relación es menor que 1, se concluye que el estimador uno es más eficiente que el dos

Utilidad de estimadores sesgados

Figure 7-6 A biased estimator $\hat{\Theta}_1$ that has smaller variance than the unbiased estimator $\hat{\Theta}_2$.



Método de Máxima Verosimilitud

- Es un método para obtener un estimador puntual de un parámetro
- Es un método genérico que puede ser aplicado a cualquier parámetro con cualquier distribución de probabilidad

Definición

- Suponga que X es una v.a. con una distribución de probabilidad $f(x;\theta)$., donde θ es un solo parámetro desconocido.
- Sean $x_1, x_2, \ldots, x_n$ los valores observados en una muestra aleatoria de tamaño n.
- La función de máxima verosimilitud de la muestra es

$$L(\theta) = f(x_1;\theta)\, f(x_2;\theta) \ldots f(x_n;\theta)$$

- Obsérvese que la función de verosimilitud es ahora función exclusiva del parámetro desconocido θ.
- *El estimador de máxima verosimilitud de θ es el valor de θ que maximiza la función de verosimilitud $L(\theta)$.*

Ejemplo: variable discreta

- Sea X una v.a de Bernoulli. La función de masa de probabilidad es

$$f(x;p)=\begin{cases} p^{x}(1-p)^{1-x}; x=0,1 \\ 0; otrocaso \end{cases}$$

- Estimar el parámetro p.

Ejemplo: variable contínua

- Sea que X tenga una distribución normal con media desconocida y varianza conocida.
- Estimar la media para una muestra aleatoria de tamaño n.

Propiedades del estimador de máxima verosimilitud

- Bajo condiciones muy generales no restrictivas, cuando el tamaño de la muestra n es grande y si $\hat{\Theta}$ es el estimador de máxima verosimilitud del parámetro θ, entonces
  - $\hat{\Theta}$ es un estimador aproximadamente insesgado
  - La varianza de $\hat{\Theta}$ es muy pequeña
  - $\hat{\Theta}$ tiene una distribución normal aproximada
- Por tanto, un estimador de máxima verosimilitud es aproximadamente un MVUE
- Para usar la estimación de máxima verosimilitud, la distribución de probabilidad debe ser conocida.

Distribuciones de muestreo

- Recordemos que un estadístico es una v.a.
- A la distribución de probabilidad de un estadístico se le llama **Distribución de muestreo**.
- La distribución de muestreo de un estadístico depende de:
  - La distribución de la población,
  - Del tamaño de la muestra
  - y del método utilizado para seleccionar la muestra

Distribuciones de muestreo de medias

- Suponga que se quiere hallar la distribución de la media muestral
- Suponga que la población tiene una distribución normal con media μ y varianza σ^2.
- Por tanto, cada observación X_i tiene una distribución normal e independiente con media μ y varianza σ^2.
- Por tanto, la media muestral será:

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

- Y tiene una distribución normal con media:

$$\mu_{\bar{x}} = \frac{\mu + \mu + \ldots + \mu}{n} = \mu$$

- Y varianza:

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2 + \sigma^2 + \ldots + \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Teorema del límite central

Ejemplo: distribución de los resultados del lanzamiento de varios dados

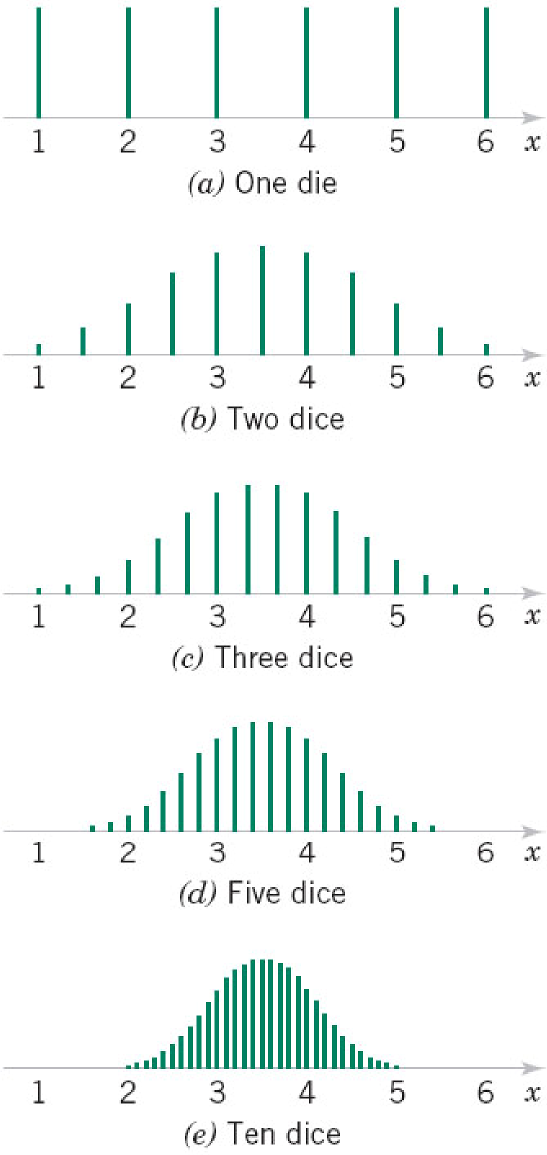


Figure 7-1
Distributions of average scores from throwing dice. [Adapted with permission from Box, Hunter, and Hunter (1978).]

Teorema del límite central

- Si X_1, X_2,…, X_n es una muestra aleatoria de tamaño n tomada de una población (sea finita o infinita) con media μ y varianza finita σ^2, y si X es la media muestral, entonces la forma límite de la distribución de

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

- Cuando n tiende a infinito, es la distribución normal estándar

En la práctica…

- Si $n \geq 30$, la aproximación normal será satisfactoria independientemente de la forma de la población.
- Si $n<30$, el teorema del límite central funcionará si la distribución de la población no se aparta significativamente de la distribución normal.

Ejemplo

- Suponga una v.a X con una distribución uniforme contínua:

$$f(x)=\begin{cases}1/2; 4 \le x \le 6\\ 0; enotrocaso\end{cases}$$

- Encuentre la distribución de la media muestral de una muestra de tamaño n=40
- Solución:
  - La media y varianza de X son μ=5 y σ^2=1/3
  - Por el teorema del límite central, para la media:

$$\mu_{\overline{x}}=\mu=5$$

$$\sigma^2_{\overline{X}}=\frac{\sigma^2}{n}=\frac{1/3}{40}=\frac{1}{120}$$

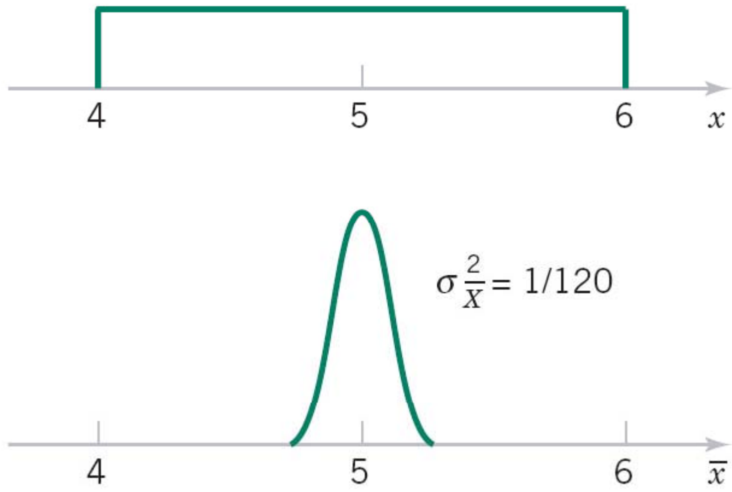


Figure 7-3 The distributions of X and $\overline{X}$ for Example 7-2.

Intervalos de Confianza

- Cuando se estima un parámetro, es necesario determinar qué tan cerca está la estimación puntual del valor real.
- Una forma de determinar la precisión de la estimación es con el error estándar
- Otra forma de estimar la precisión es con los intervalos de confianza

Intervalos de Confianza

- Se puede determinar que el valor desconocido θ está en un intervalo $l \leq \theta \leq u$
- Los valores de los límites dependen del valor numérico del estadístico para una muestra particular
- Diferentes muestras producen diferentes valores del estadístico y de los límites del intervalo.

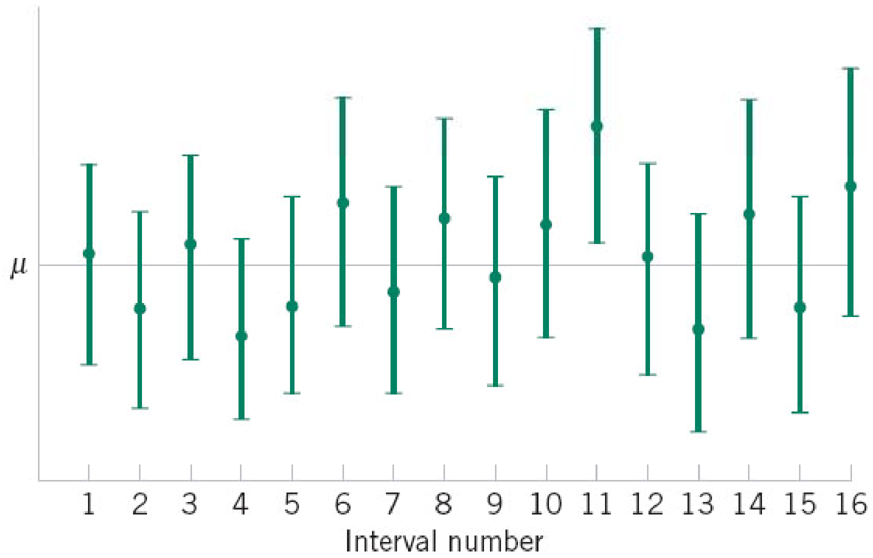


Figure 8-1 Repeated construction of a confidence interval for μ.

Intervalos de confianza

- L y U son variables aleatorias que representan los límites superior e inferior de los intervalos de confianza
- Pueden determinarse unos valores de L y U de manera que:

$$P(L \leq \theta \leq U) = 1 - \alpha$$

- Donde $0 < \alpha < 1$.
- Por tanto, se tendrá una probabilidad $1-\alpha$ de seleccionar una muestra que producirá un intervalo que incluya el valor verdadero de θ.

Intervalos de confianza

- Al intervalo que resulta:

$$l \leq \theta \leq u$$

Se le llama un **intervalo de confianza** del 100(1-α) por ciento para el parámetro θ.

- A las cantidades l y u se les llama **límites de confianza inferior y superior**, respectivamente.
- A (1-α) se le llama **coeficiente de confianza**.
- Una forma de calcular los límites inferior y superior es sumar y restar respectivamente un múltiplo del error estándar al valor estimado.

Interpretación

- Si se toma un número infinito de muestras aleatorias y se calcula un intervalo de confianza del 100(1-α) por ciento para θ en cada muestra, entonces el 100(1-α) por ciento de estos intervalos incluirán el valor real de θ.

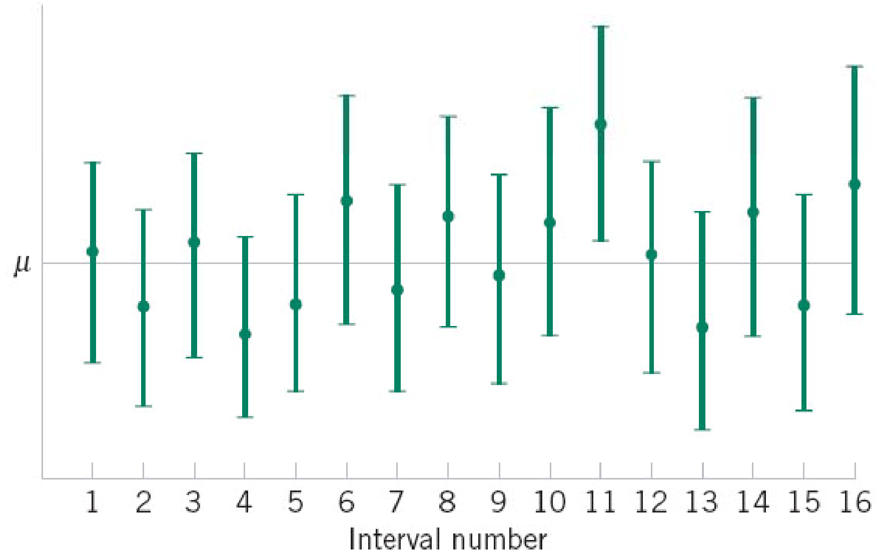


Figure 8-1 Repeated construction of a confidence interval for μ.