# Intervalos de confianza

Contenidos adaptados del libro "Probability and statistics for computer scientists, Second edition, M. Baron" (Capítulo 9)

# Contenidos

# 1. Objetivos

- Comprender la importancia de la estimación
- Reconocer situaciones en que los diferentes métodos de muestreo son relevantes
- Diferenciar entre parámetro de una población y estadístico de una muestra
- Diferenciar entre estimadores sesgados e insesgados
- Diferenciar entre la estimación puntual y la estimación por intervalo
- Construir intervalos de confianza para la media, varianza proporción de una población
- Construir intervalos de confianza para la diferencia de medias y proporciones de dos poblaciones

# 2. Introducción a la inferencia estadística

- La inferencia estadística consiste en estimar (inferir) los parámetros (media, desviación estándar, ..) de una población (personas, cosas, ..) a partir de una parte de la misma.
- En inferencia estadística se trabaja con incertidumbre (variables aleatorias) al no disponer de todos los datos de la población, por lo que se manejan distribuciones de probabilidad para minimizar el error que se produce al realizar una estimación
- El resultado de la inferencia estadística puede ser
  - Un intervalo de confianza o conjunto de posibles valores para los parámetros estimados, junto con una probabilidad de error cometido
  - La comprobación de una hipótesis sobre los valores para los parámetros estimados, junto con una probabilidad de error cometido
  - Un modelo de predicción (regresión) del valor de una propiedad de un elemento de la población a partir del valor de otra propiedad.

# 3. Muestreo Definiciones (I)

- Población
  - Todas las unidades o elementos de interés sobre cuyas propiedades (peso, tamaño, etc.) se quiere realizar un estudio estadístico (descriptivo o inferencial). Pueden ser personas, animales, planetas, cosas, ..
- Muestra (*Sample*)
  - Subconjunto de una población compuesto por una parte de los elementos de la población (denominadas observaciones).
- Propiedad (de cada elemento de una población)
  - Dato que representa alguna medida sobre un elemento de una población. Por ejemplo, la edad de cada persona en una población de personas. Es lo que en estadística descriptiva se denomina “variable estadística”.
- Muestreo (*Sampling*)
  - Acción de creación de una muestra y recolección de los valores de las propiedades de los elementos incluidos en la muestra.
- Parámetro (de una población)
  - Cualquier característica numérica de una población. Se utiliza para hacer afirmaciones sobre la población. Por ejemplo, media, mediana, moda, varianza, coeficiente de correlación entre dos propiedades.
- Estadístico (de una muestra)
  - Cualquier característica numérica de una muestra. Un estadístico se utiliza como estimador del posible valor de un parámetro. Por ejemplo, la media.

# 3. Muestreo
# Definiciones (II)

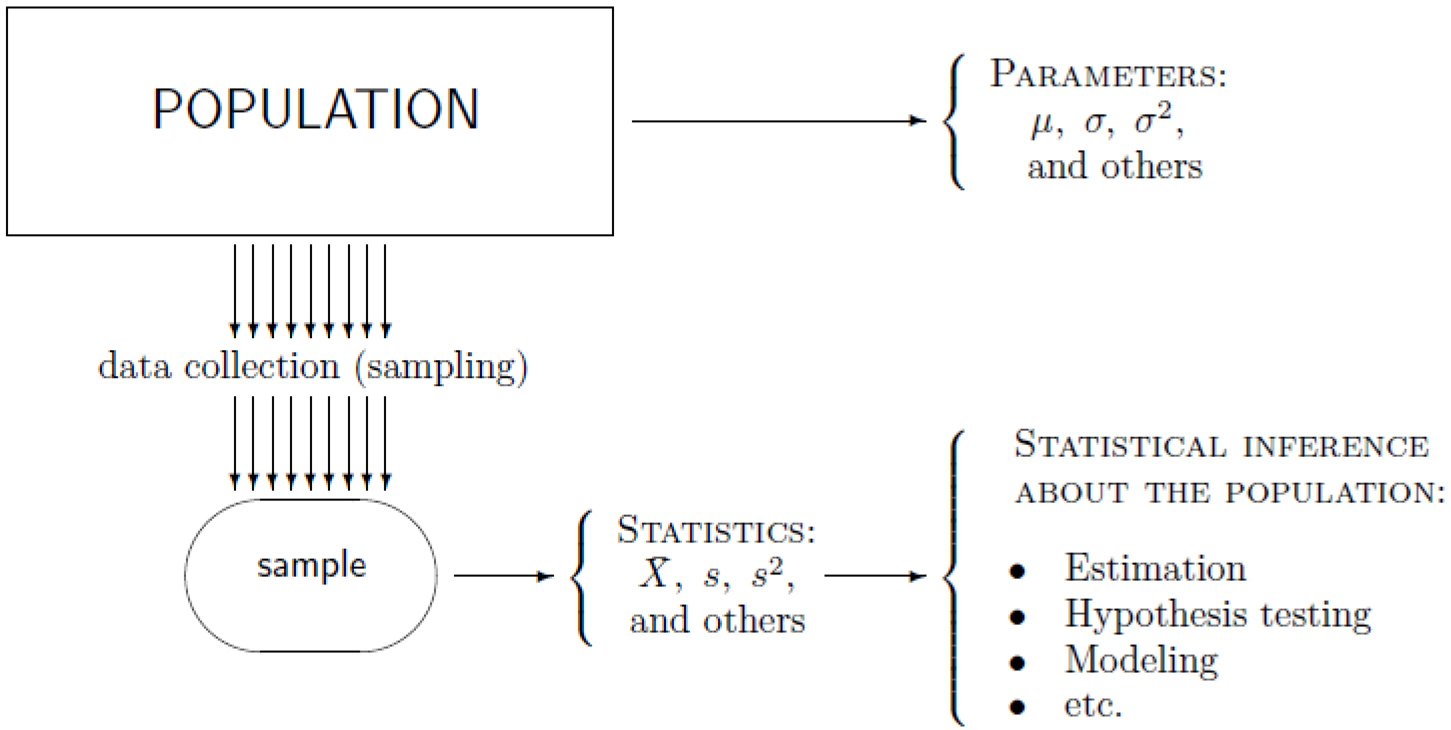



FIGURE 8.1: *Population parameters and sample statistics.*

# 3. Muestreo
# Métodos de muestreo

- Para garantizar que una muestra sea representativa de una población y no esté sesgada, se pueden aplicar diferentes métodos de muestreo
- Muestreo aleatorio simple
  - Cuando los elementos se recogen de toda la población independientemente uno de otro, teniendo cada elemento la misma probabilidad de ser elegido.
- Muestreo aleatorio sistemático
  - Se elige un elemento y a partir de él se selecciona el resto aplicando un procedimiento sistemático. Ejemplo: enumerando todos los elementos y, a partir del elegido se seleccionan los que se encuentren a cada 10 posiciones de él.
- Muestreo aleatorio estratificado
  - Se divide la población en estratos o categorías con el fin de garantizar que todos los estratos de interés están representados en la muestra. Ejemplo: sondeo de opinión.
  - La distribución de la muestra en función de los diferentes estratos se denomina afijación, y puede ser de diferentes tipos:
    - Afijación Simple: A cada estrato le corresponde igual número de elementos
    - Afijación Proporcional: Se toman elementos proporcionalmente al tamaño de cada estrato
- Muestreo aleatorio por conglomerados
  - Se divide la población en conglomerados o grupos compactos (por ejemplo, áreas geográficas), se eligen aleatoriamente varios conglomerados, y se hace un muestreo aleatorio en cada uno de ellos.
- Muestreo no aleatorio
  - Cuando el muestreo está condicionado por alguna causa como, por ejemplo, limitarlo a personas que cumplen una condición sobre su edad y estado civil.

# 3. Muestreo Reemplazamiento

- Muestreo con reemplazamiento es aquel en el que cada elemento de una población puede ser seleccionado más de una vez
  - Todos los elementos de la población tienen la misma probabilidad siempre
- Muestreo sin remplazamiento es aquel en el que cada elemento sólo puede ser seleccionado una sola vez
  - Una vez elegido un elemento de la población para incluirlo en la muestra, se excluye de la población para que no pueda volver a ser elegido.
- En poblaciones grandes no suele haber importantes diferencias al realizar inferencia estadística, pero sí en poblaciones pequeñas

# 4. Estimación de parámetros

- Un estimador $\hat{\theta}$ es una función aplicada a una muestra de una población para obtener un posible valor para un parámetro $\theta$ de la población.
  - **Un estimador $\hat{\boldsymbol{\theta}}$ de un parámetro es una variable aleatoria**, su valor puede variar de una muestra a otra de la misma población
  - **Un parámetro $\boldsymbol{\theta}$ no es una variable aleatoria**, es una constante con un solo valor para una población
- Como estimadores de parámetros se suelen utilizar estadísticos con el mismo nombre, pero diferente "apellido"
  - Parámetros: Media poblacional, Desviación estándar poblacional, Varianza poblacional, Proporción poblacional, Cuantil poblacional, Mediana poblacional, …
  - Estadísticos: Media muestral, Desviación estándar muestral, Varianza muestral, Proporción muestral, Cuantil muestral, Mediana muestral, …
- Propiedades de un estimador
  - Sesgo, consistencia y normalidad asintótica

# 4. Estimación de parámetros Estimador insesgado (*unbiased*)

- Un estimador $\hat{\theta}$ es insesgado para un parámetro $\theta$, si la esperanza del estimador es igual al valor del parámetro para todos los posibles valores del parámetro.
  - $\boldsymbol{E(\hat{\theta}) = \theta}$
  - Ejemplo: Media muestral como estimador de la media poblacional
- El sesgo (*bias*) de un estimador $\hat{\theta}$ es la esperanza de la diferencia (error) entre su valor y el valor real del parámetro que estima
  - $Bias(\hat{\theta}) = E(\hat{\theta} - \theta)$
  - Ejemplo: El sesgo de la media muestral es cero como estimador de la media poblacional

# 4. Estimación de parámetros
# Estimador consistente

- Un estimador $\hat{\theta}$ es consistente para un parámetro $\theta$, si la probabilidad de su error muestral (diferencia entre su valor y el valor real del parámetro) **tiende a cero cuando el tamaño de la muestra tiende a infinito**.
  - $\mathrm{P}\left(\left|\hat{\theta}-\theta\right| > \varepsilon\right) \to 0 \text{ cuando } n \to \infty, \quad \text{donde } \varepsilon > 0$
- Es decir, cuando estimamos $\theta$ a partir de una muestra grande, es poco probable que el error de la estimación $\left|\hat{\theta}-\theta\right|$ sea mayor que cualquier valor $\varepsilon$, y lo hace con probabilidades cada vez más pequeñas a medida que aumentamos el tamaño de la muestra.

# 4. Estimación de parámetros
# Estimador asintóticamente normal

- Un estimador es una variable aleatoria
- **La distribución de probabilidad de un estimador converge hacia una distribución Normal cuando el tamaño de la muestra aumenta**.
- Se dice que el estimador es asintóticamente normal
  - Es la aplicación del Teorema Central del Límite
- Si el estimador se estandariza y se convierte en una variable Z, ésta converge a una distribución Normal estándar N(0,1)
  - $Z = \frac{\hat{\theta} - E(\hat{\theta})}{\sqrt{Var(\hat{\theta})}}$

# 4. Estimación de parámetros
# Estimación puntual o por intervalo

- Si se calcula un único valor de un estimador, se está realizando una **estimación puntual** (o por punto).
  - Habría que estimar también el posible error cometido
- Si se calcula un rango de valores posibles para un parámetro se está realizando una **estimación por intervalo**
  - Es habitual en inferencia estadística
  - Realmente se utilizan dos estimadores: uno para el límite inferior del intervalo y otro para el límite superior
  - Se suele denominar "intervalo de confianza", porque además de los extremos del intervalo, se calcula el nivel de confianza o probabilidad de que el valor del parámetro se encuentre en dicho intervalo.

# 4. Estimación de parámetros Cálculo de los estimadores

- Como estimadores de los parámetros de una población se suelen utilizar los estadísticos de una muestra de esa población
- Por lo que es importante saber calcular los estadísticos
- En inferencia se suelen usar tres estadísticos: Media, Varianza y Proporción.

| **Estimador (Estadístico)** | **Parámetro que estima** |
|---|---|
| Media muestral ($\bar{X}$) | Media poblacional ($\mu$) |
| Varianza muestral ($S^2$) | Varianza poblacional ($\sigma^2$) |
| Proporción muestral ($\hat{P}$) | Proporción poblacional ($p$) |

# 4. Estimación de parámetros Cálculo del estimador: media muestral

- Muestra de valores de una propiedad para n elementos de la población
  - Ejemplo: Los pesos de n personas
  - $X = (X_1, X_2, \ldots, X_n)$
- Parámetro a estimar: media de la propiedad para toda la población (ej. peso medio)
  - $\theta = \mu = \frac{x_1 + x_2 + \cdots + x_{tamaño\ total\ de\ la\ población}}{tamaño\ total\ de\ la\ población}$
- Estimador
  - $\hat{\theta} = \bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$
- Es un estimador insesgado
  - $E(\hat{\theta}) = E(\bar{X}) = \mu = \theta$
- Varianza del estimador
  - $Var(\hat{\theta}) = Var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{Var(X_1) + Var(X_2) + \cdots + Var(X_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$
- Es un estimador consistente
  - Su error (varianza) tiende a cero cuanto mayor es el tamaño de la muestra
- Es un estimador asintóticamente normal
  - Tiene una distribución aproximadamente Normal cuando n es grande

# 4. Estimación de parámetros
# Cálculo del estimador: varianza muestral

- Muestra de valores de una propiedad para n elementos de la población
  - Ejemplo: Los pesos de n personas
  - $X = (X_1, X_2, \dots, X_n)$
- Parámetro poblacional a estimar
  - $\theta = \sigma^2 = \frac{1}{tamaño\ total\ de\ la\ población} \sum_{i=1}^{tamaño\ total\ de\ la\ población} (x_i - \mu)^2$
- Estimador
  - $\hat{\theta} = s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$
- Es un estimador insesgado
  - $E(\hat{\theta}) = E(s^2) = \sigma^2 = \theta$ Lo asegura el coeficiente $\frac{1}{n-1}$
- Es un estimador consistente
  - Su error tiende a cero cuanto mayor es el tamaño de la muestra
- Es un estimador asintóticamente normal
  - Tiene una distribución aproximadamente Normal cuando n es grande

# 4. Estimación de parámetros
# Cálculo del estimador: proporción muestral

- Muestra de valores de una propiedad para n elementos de la población
  - Ejemplo: Los pesos de n personas
  - $X = (X_1, X_2, \ldots, X_n)$
- Parámetro poblacional a estimar
  - $\theta = p = \dfrac{\text{número de elementos de la población que cumplen una condición } C}{\text{tamaño total de la población}}$
- Estimador
  - $\hat{\theta} = \hat{P} = \dfrac{\text{número de elementos de la muestra que cumplen la condición } C}{n}$
- Es un estimador insesgado
  - $E(\hat{\theta}) = E(\hat{P}) = p = \theta$
- Es un estimador consistente
  - Su error tiende a cero cuanto mayor es el tamaño de la muestra
- Es un estimador asintóticamente normal
  - Tiene una distribución aproximadamente Normal cuando n es grande, porque es similar a una media muestral

# 5. Definición de Intervalo de Confianza

- Un intervalo $[a, b]$ es un intervalo de confianza para un parámetro $\theta$ de una población, con un nivel de confianza de $(1-\alpha)$, si el intervalo contiene el valor del parámetro con una probabilidad de $(1-\alpha)$.
  - $P\{a \leq \theta \leq b\} = 1-\alpha$
- **El nivel de confianza $(\mathbf{1}-\boldsymbol{\alpha})$ o probabilidad de cobertura, es un valor entre 0 y 1, y representa la probabilidad de acertar en la predicción**, es decir la probabilidad de que el valor real del parámetro se encuentre entre los límites del intervalo.
- El **nivel de error o nivel de significación $\boldsymbol{\alpha}$**, representa lo contrario que la confianza, es decir, es la probabilidad de no acertar en la predicción.
  - $P\{(\theta < a) \cup (\theta > b\} = \alpha$

# 5. Definición de Intervalo de Confianza Intervalos y muestras

- Un intervalo de confianza se construye a partir de una muestra de la población
  - Para una misma población puede haber intervalos de confianza diferentes según la muestra elegida
- Recordar que
  - Un parámetro $\theta$ de una población (por ejemplo, la media poblacional) no es una variable aleatoria, es un valor constante
  - Un estimador $\hat{\theta}$ de un parámetro (por ejemplo, la media muestral) sí es una variable aleatoria

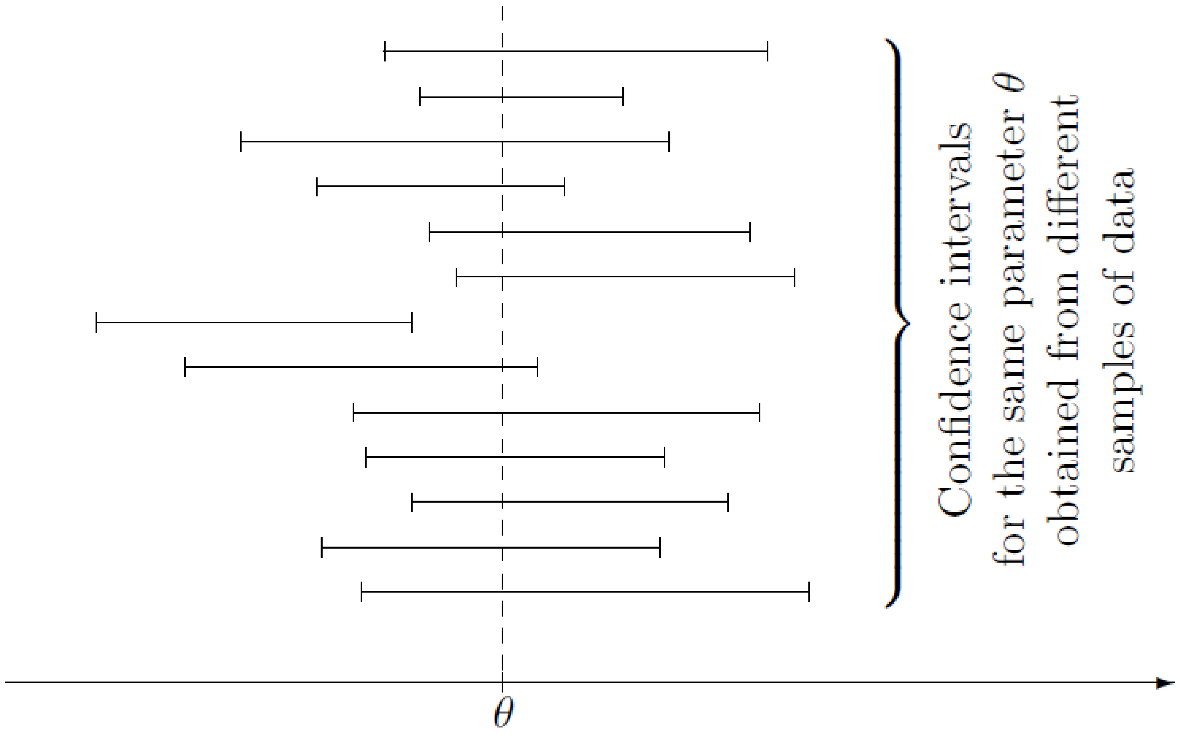



FIGURE 9.2: *Confidence intervals and coverage of parameter $\theta$.*

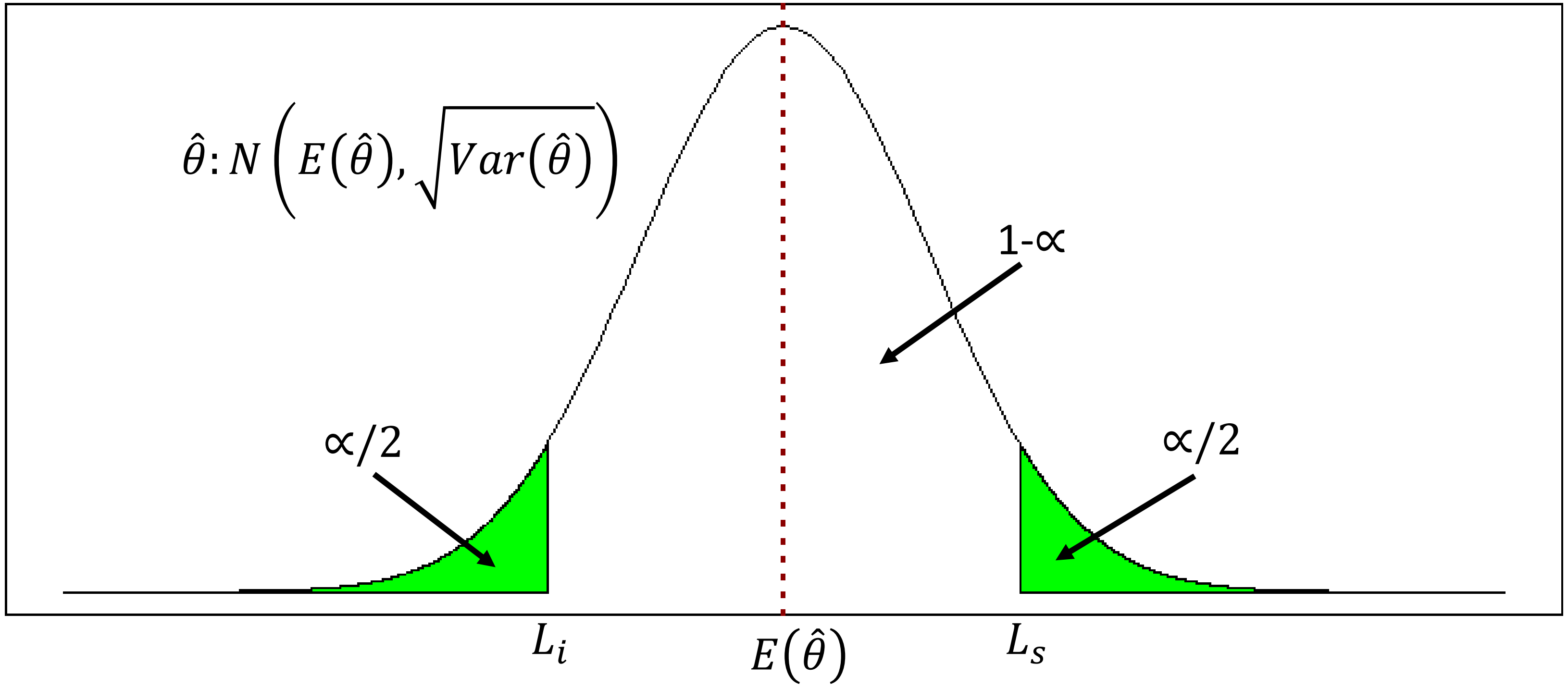
# 5. Definición de Intervalo de Confianza Ejemplo

- En una población de personas, una propiedad de cada persona es su peso
- Un parámetro es la media del peso de la población
- Si el intervalo de confianza para para la media poblacional del peso es $[33.55, 66.45]$ y $(1-\alpha)$=0.90
- Significa que
  - La probabilidad de acertar, es decir, de que el peso medio de la población esté entre 33.55kg y 66.45kg es del 90%
  - El 90% se refiere a la proporción de intervalos de confianza que contienen el peso medio de la población
  - Es decir, si calculásemos 100 intervalos de confianza para el peso medio de la población, haciendo 100 muestreos diferentes, en 90 de ellos estará el rango $[33.55, 66.45]$ y en 10 de ellos no.

# 6. Método general de construcción de un intervalo de confianza

- 1º) Se obtiene una muestra de la población
- 2º) Se calcula el valor del estimador $\hat{\theta}$ para el parámetro $\theta$
  - Se utiliza un estimador insesgado
- 3º) Se calcula el margen $\varepsilon$ que hay que sumar y restar al estimador para obtener los límites del intervalo
  - Para calcular el margen, se estandariza el estimador asumiendo que tiene distribución de probabilidad Normal
- 4º) Se calculan los límites inferior y superior del intervalo
  - $L_i = \hat{\theta} - \varepsilon$
  - $L_s = \hat{\theta} + \varepsilon$
  - El intervalo se suele denotar como $\hat{\theta} \pm \varepsilon$ o como $\left[\hat{\theta} - \varepsilon, \hat{\theta} + \varepsilon\right]$

# 6. Método general
# Interpretación gráfica sin estandarizar




$L_i$ = cuantil $\propto/2$ de la variable $\hat{\theta}: N\left(E(\hat{\theta}), \sqrt{Var(\hat{\theta})}\right)$

Porque $P(\hat{\theta}<L_i) = \propto/2$

$L_s$ = cuantil $1-\propto/2$ de la variable $\hat{\theta}: N\left(E(\hat{\theta}), \sqrt{Var(\hat{\theta})}\right)$

Porque $P(\hat{\theta}<L_s) = 1-\propto/2$

# 6. Método general
# Estandarización del estimador

- El estimador $\hat{\theta}$ se convierte en una variable $Z: N(0,1)$
  - $Z = \frac{\hat{\theta}-E(\hat{\theta})}{\sigma(\hat{\theta})} = \frac{\hat{\theta}-\theta}{\sigma(\hat{\theta})}$ al ser un estimador insesgado $E(\hat{\theta}) = \theta$
- Los valores de la variable estandarizada $Z$ correspondientes a los cuantiles (q): $\alpha/2$ y $(1-\alpha/2)$ se denominan valores críticos, y se denotan como
  - $-z_{\alpha/2} = q(\alpha/2) \rightarrow P\{Z < -z_{\alpha/2}\} = \frac{\alpha}{2}$
  - $z_{\alpha/2} = q(1-\alpha/2) \rightarrow P\{Z < z_{\alpha/2}\} = 1 - \frac{\alpha}{2}$
- Por lo que
  - $P\left\{-z_{\alpha/2} \leq \frac{\hat{\theta}-E(\hat{\theta})}{\sigma(\hat{\theta})} \leq z_{\alpha/2}\right\} = (1-\alpha) \rightarrow P\{\hat{\theta} - z_{\alpha/2} \cdot \sigma(\hat{\theta}) \leq \theta \leq \hat{\theta} + z_{\alpha/2} \cdot \sigma(\hat{\theta})\} = (1-\alpha)$
- Y el intervalo de confianza es
  - $[\hat{\theta} - z_{\alpha/2} \cdot \sigma(\hat{\theta}),\ \ \hat{\theta} + z_{\alpha/2} \cdot \sigma(\hat{\theta})]$

- Nota: Recordar el concepto de <u>cuantil</u>
  - *Dada una variable aleatoria X, el cuantil q, es el valor x de la variable, tal que* $P\{X \leq x\} = q$
  - En una variable aleatoria continua $P\{X \leq x\} = P\{X < x\} = q$
  - Un cuantil q equivale a un percentil qx100

# 6. Método general
# Interpretación gráfica con estandarización

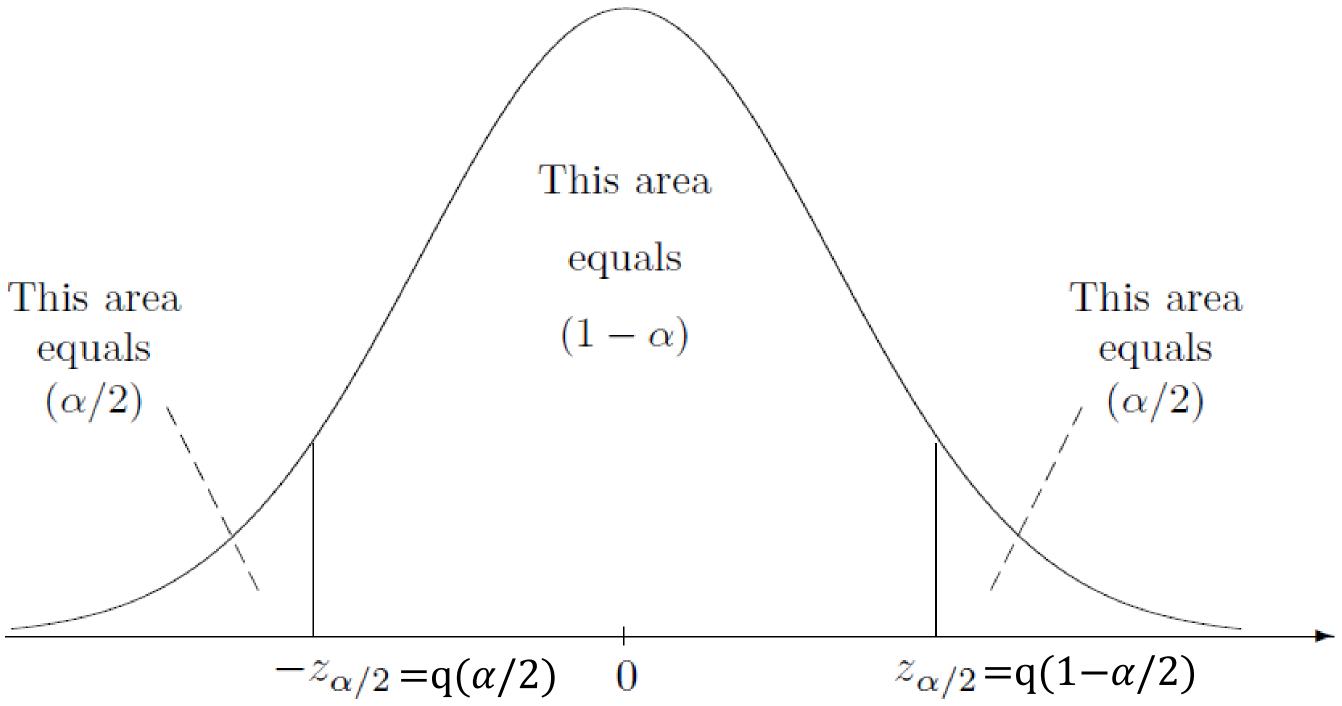



FIGURE 9.3: *Standard Normal quantiles $\pm z_{\alpha/2}$ and partition of the area under the density curve.*

# 6. Método general Estandarización del estimador. Valores típicos

- En los trabajos de inferencia estadística se utilizan habitualmente unos pocos valores de α, el más habitual es 0.05, correspondiente a un nivel de confianza de 0.95 o 95%

| $\boldsymbol{1-\alpha}$ | $\boldsymbol{\alpha}$ | $\boldsymbol{z_{\alpha/2}}$ |
|---|---|---|
| 0.90 | 0.10 | 1.64 |
| 0.95 | 0.05 | 1.96 |
| 0.99 | 0.01 | 2.58 |

# 7. Intervalo de confianza para la media poblacional

- Casos posibles
  - Varianza poblacional conocida → Se utiliza un estimador con distribución Normal
  - Varianza poblacional desconocida
    - Muestra grande → Se utiliza un estimador con distribución Normal
    - Muestra pequeña, con distribución Normal de los datos de la población → Se utiliza un estimador con distribución T de Student
    - Muestra pequeña, sin distribución Normal de la población o desconocida → Se aplican técnicas de inferencia no paramétrica (ver capítulo 10.2 del libro)
- Notas
  - Una muestra se suele considerar grande cuando el tamaño es $n \geq 30$ o $n > 30$ (no hay unanimidad)
  - La condición de normalidad se aplica a la población, si no se dispone de esa información, puede comprobarse <u>la normalidad de la muestra</u>, aunque no sea del todo fiable, combinando métodos analíticos y métodos gráficos

# 7. Intervalo de confianza para la media poblacional. Fórmulas

| Se conoce la varianza poblacional ($\sigma^2$) | Tamaño de la muestra (n) | Intervalo de confianza |
|---|---|---|
| Si | Cualquiera | $L_i = \bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ <br> $L_s = \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ |
| No | Grande | $L_i = \bar{x} - z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$ <br> $L_s = \bar{x} + z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$ |
| No | Pequeña con distribución Normal | $L_i = \bar{x} - t_{\frac{\alpha}{2},n-1} \frac{s}{\sqrt{n}}$ <br> $L_s = \bar{x} + t_{\frac{\alpha}{2},n-1} \frac{s}{\sqrt{n}}$ |

# 7. Intervalo de confianza para la media poblacional. Valores para el cálculo

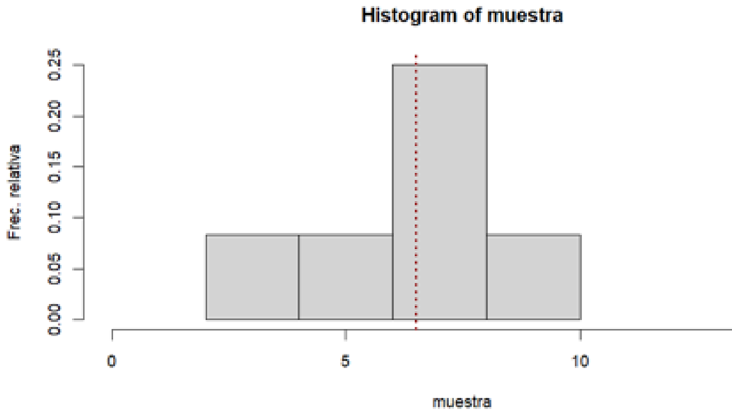
- Parámetro $\theta$ = Media poblacional = $\mu$
- Valores necesarios para el cálculo
  - Estimador $\hat{\theta}$ = Media muestral = $\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n}$
  - Tamaño de la muestra = $n$
  - Varianza poblacional = $\sigma^2$
    - Desviación estándar poblacional = $\sigma$
  - Varianza muestral = $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
    - Desviación estándar muestral = s
  - Nivel de error o significación = $\alpha$
    - Nivel de confianza = $1 - \alpha$
  - Valor crítico para muestra grande (el estimador se considera Normal estándar N(0,1)
    - $z_{\frac{\alpha}{2}} = q\left(1 - \frac{\alpha}{2}\right)$ → Se obtiene con la tabla A4 o con software
  - Valor crítico para muestra pequeña (el estimador se considera T de Student con n-1 grados de libertad
    - $t_{\frac{\alpha}{2},n-1} = q\left(1 - \frac{\alpha}{2}\right)$ → Se obtiene con la tabla A5 o con software

# 7. Intervalo de confianza para la media poblacional Ejemplo 9.13a (Mediciones)

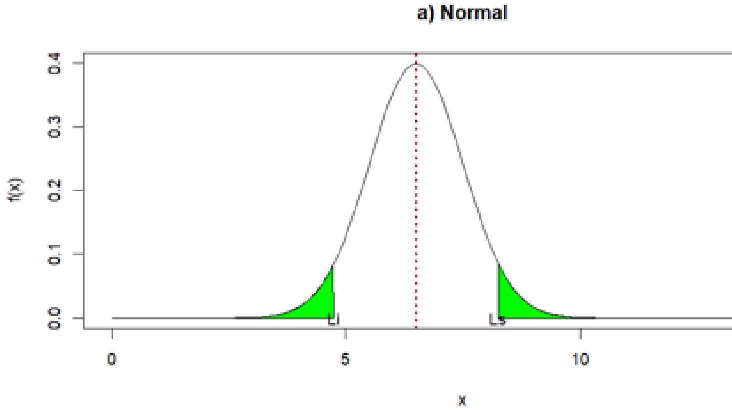
- Construir un intervalo de confianza del 95 % para la media poblacional basado en una muestra de mediciones con distribución Normal
  - 2.5, 7.4, 8.0, 4.5, 7.4, 9.2
- En los casos:
  a) **Se sabe que el dispositivo de medición garantiza una desviación estándar de $\sigma$ = 2.2**
  b) Se desconoce el valor de la desviación estándar

# 7. Intervalo de confianza para la media poblacional Ejemplo 9.13 (Solución a)

- Estimador $\hat{\theta}$ = Media muestral = $\bar{X}$
- Valor del estimador para la muestra = $\bar{x} = \frac{2.5+7.4+8.0+4.5+7.4+9.2}{6} = 6.50$
- Tamaño de la muestra = $n = 6$
- Desviación estándar poblacional = $\sigma = 2.2$
- Nivel de significación = $\alpha = 1 - 0.95 = 0.05$
- Es una muestra pequeña n<30, pero con distribución Normal
- $z_{\alpha/2} = q(1 - \alpha/2) = q(1 - 0.025) = q(0.975) = 1.96$
    - Se obtiene con la tabla A4 o con software como R




a) Intervalo de confianza con varianza conocida
- $L_i = \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 6.50 - (1.96)\frac{2.2}{\sqrt{6}} = 4.74$
- $L_s = \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 6.50 + (1.96)\frac{2.2}{\sqrt{6}} = 8.26$



# 7. Intervalo de confianza para la media poblacional Muestra grande y varianza desconocida

- Se asume una distribución Normal para el estimador $\bar{X}$: $N\left(E(\bar{X}), \sigma(\bar{X})\right) = N(\mu, s/\sqrt{n})$
- El estimador $\bar{X}$ se estandariza, convirtiéndolo en una variable $Z$: $N(0,1)$
  - $Z = \frac{\bar{X} - E(\bar{X})}{\sigma(\bar{X})} = \frac{\bar{X} - \mu}{s/\sqrt{n}}$
    - $E(\bar{X}) = \mu$ al ser un estimador insesgado
    - $\sigma(\bar{X}) = s(\bar{X}) = \frac{s}{\sqrt{n}}$ al desconocerse la varianza poblacional, se utiliza en su lugar la varianza muestral $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$
- Los valores de la variable estandarizada $Z$ correspondientes a los cuantiles (q): $\frac{\alpha}{2}$ y $\left(1 - \frac{\alpha}{2}\right)$ se denominan valores críticos, y se denotan como
  - $-z_{\frac{\alpha}{2}} = q\left(\frac{\alpha}{2}\right) \rightarrow P\left\{Z < -z_{\frac{\alpha}{2}}\right\} = \frac{\alpha}{2}$
  - $z_{\frac{\alpha}{2}} = q\left(1 - \frac{\alpha}{2}\right) \rightarrow P\left\{Z < z_{\frac{\alpha}{2}}\right\} = 1 - \frac{\alpha}{2}$
- Por lo que
  - $P\left\{-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{s/\sqrt{n}} \leq z_{\frac{\alpha}{2}}\right\} = (1 - \alpha) \rightarrow P\left\{\bar{X} - z_{\frac{\alpha}{2}} \cdot s/\sqrt{n} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \cdot s/\sqrt{n}\right\} = (1 - \alpha)$
- Y el intervalo de confianza utilizando una muestra concreta con media $\bar{x}$ y desviación $s$ es
  - $\left[\bar{x} - z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\ , \bar{x} + z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\ \right]$

# 7. Intervalo de confianza para la media poblacional Muestra pequeña y varianza desconocida

- Cuando la muestra es pequeña, en lugar se asumir que el estimador tiene una distribución Normal, se considera que tiene una distribución T de Student con n-1 grados de libertad
  - Se obtiene un intervalo más ancho
- El estimador $\hat{\theta}$ se convierte en una variable $T(n-1)$
  - $T(n-1) = \frac{\bar{X}-E(\bar{X})}{s(\bar{X})} = \frac{\bar{X}-\mu}{s/\sqrt{n}}$
- Los valores de la variable T correspondientes a los cuantiles $\frac{\alpha}{2}$ y $1-\frac{\alpha}{2}$ se denominan valores críticos, y se denotan como
  - $-t_{\frac{\alpha}{2},n-1} = q\left(\frac{\alpha}{2}\right) \to P\left\{T < -t_{\frac{\alpha}{2},n-1}\right\} = \frac{\alpha}{2}$
  - $t_{\frac{\alpha}{2},n-1} = q\left(1-\frac{\alpha}{2}\right) \to P\left\{T < t_{\frac{\alpha}{2},n-1}\right\} = 1-\frac{\alpha}{2}$
- Por lo que
  - $P\left\{-t_{\frac{\alpha}{2},n-1} \leq \frac{\bar{X}-\mu}{s/\sqrt{n}} \leq t_{\frac{\alpha}{2},n-1}\right\} = (1-\alpha) \to P\left\{\bar{X} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}} \leq \bar{X} \leq \bar{X} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}\right\} = (1-\alpha)$
- Y el intervalo de confianza utilizando una muestra concreta con media $\bar{x}$ y desviación $s$ es es
  - $\left[\bar{x} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}\,, \bar{x} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}\ \right]$

# 7. Intervalo de confianza para la media poblacional
## Muestra pequeña y varianza desconocida (Tabla T de Student)

- Para obtener $t_{\frac{\alpha}{2},n-1}$ se puede usar la tabla A5
- Ejemplo
  - $\alpha/2 = 0.005$
  - Grados de libertad: $n - 1 = 3$
- Hay que localizar la fila 3
- Hay que localizar la columna 0.005
- En la celda correspondiente se encuentra el valor de $t_{0.005}$ buscado = 5.84

**Table A5. Table of Student's T-distribution**

$t_\alpha$; critical values, such that $P\{t > t_\alpha\} = \alpha$

| $\nu$ | $\alpha$, the right-tail probability | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| (d.f.) | .10 | .05 | .025 | .02 | .01 | .005 | .0025 | .001 | .0005 | .0001 |
| 1 | 3.078 | 6.314 | 12.706 | 15.89 | 31.82 | 63.66 | 127.3 | 318.3 | 636.6 | 3185 |
| 2 | 1.886 | 2.920 | 4.303 | 4.849 | 6.965 | 9.925 | 14.09 | 22.33 | 31.60 | 70.71 |
| 3 | 1.638 | 2.353 | 3.182 | 3.482 | 4.541 | 5.841 | 7.453 | 10.21 | 12.92 | 22.20 |
| 4 | 1.533 | 2.132 | 2.776 | 2.999 | 3.747 | 4.604 | 5.598 | 7.173 | 8.610 | 13.04 |

# 7. Intervalo de confianza para la media poblacional Ejemplo 9.13b (Mediciones)

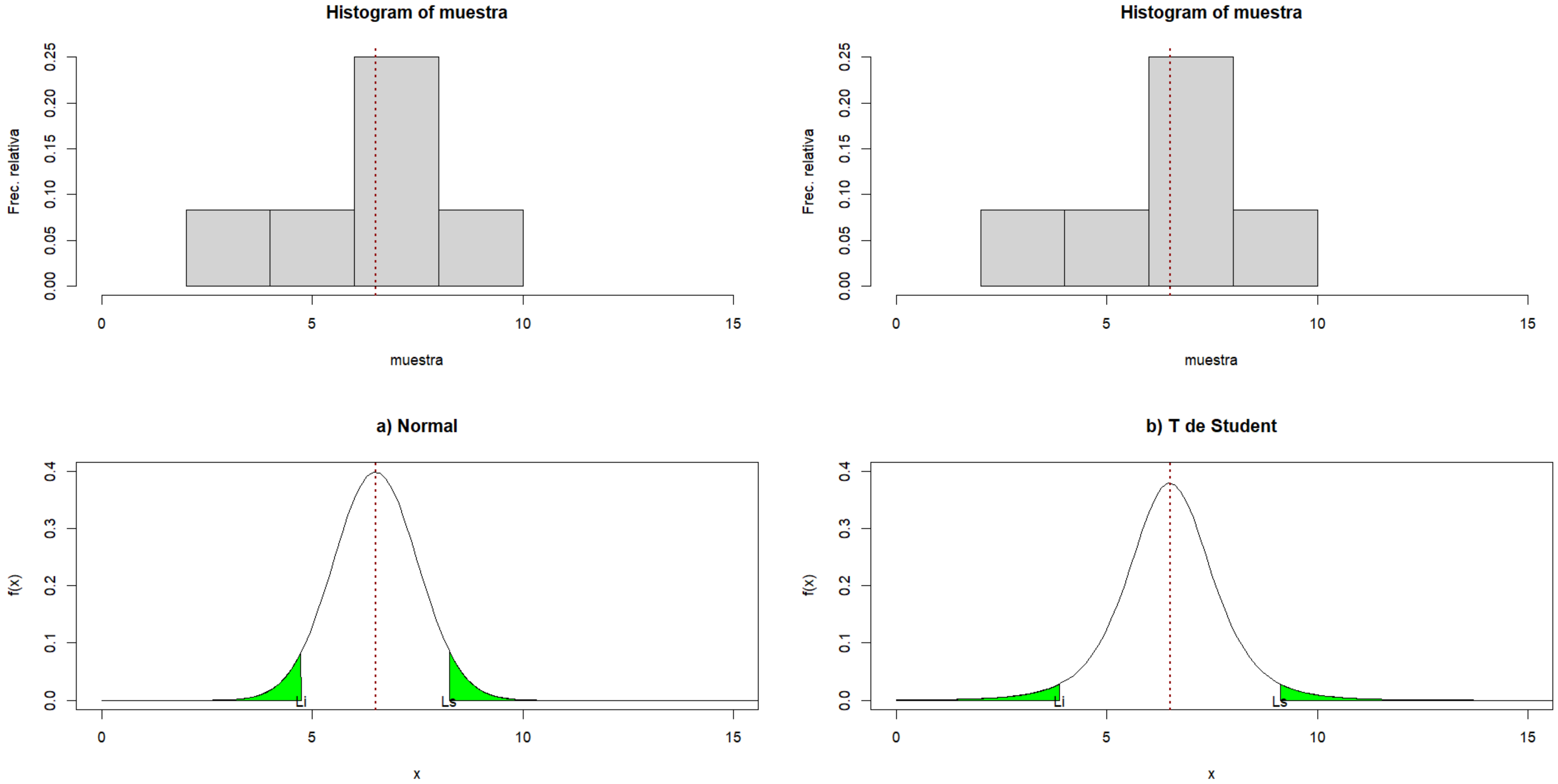
- Construir un intervalo de confianza del 95 % para la media poblacional basado en una muestra de mediciones con distribución Normal
  - 2.5, 7.4, 8.0, 4.5, 7.4, 9.2
- En los casos:
  a) Se sabe que el dispositivo de medición garantiza una desviación estándar de $\sigma$ = 2.2
  **b) Se desconoce el valor de la desviación estándar**

# 7. Intervalo de confianza para la media poblacional Ejemplo 9.13 (Solución b)

b) Intervalo de confianza con varianza desconocida y muestra pequeña

- Desviación estándar muestral = s
  - $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{6-1}\left((2.5-6.5)^2 + (7.4-6.5)^2 + (8.0-6.5)^2 + (4.5-6.5)^2 + (7.4-6.5)^2 + (9.2-6.5)^2\right) = 6.23$
  - $s = \sqrt{6.23} = 2.5$
  - $t_{\alpha/2,n-1} = t_{0.025,5} = 2.57$
    - Usando un software como R o buscando en la tabla A5, fila 5 y columna 0.025
- Intervalo de confianza
  - $L_i = \bar{x} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}} = 6.50 - (2.57)\frac{2.5}{\sqrt{6}} = 3.88$
  - $L_s = \bar{x} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}} = 6.50 + (2.57)\frac{2.5}{\sqrt{6}} = 9.12$

# 7. Intervalo de confianza para la media poblacional
# Ejemplo 9.13 (Interpretación gráfica)



# 7. Intervalo de confianza para la media poblacional Ejercicios propuestos

- Ejercicios 9.7(a), 9.8, 9.9(a) y 9.12 del libro
  - Las respuestas de 9.7 y 9.9 están disponibles en el libro

# 8. Intervalo de confianza para una proporción poblacional

- Parámetro $\theta$ = Proporción poblacional = $p$
  - $p = \frac{\textit{número de elementos de la población que cumplen una condición } C}{\textit{tamaño total de la población}}$
  - Ejemplo:
    - En una población de personas, la proporción de personas que saben conducir es del 60%
    - En este caso, $p = 0.6$
- Se aplica sólo a muestras grandes
  - Una muestra se suele considerar grande cuando el tamaño es n≥30 o n>30 (no hay unanimidad).

# 8. Intervalo de confianza para una proporción poblacional. Estimador

- Valores necesarios para el cálculo
  - Estimador $\hat{\theta}$ = Proporción muestral = $\hat{P}$
    - $\hat{P} = \frac{Número\ de\ elementos\ de\ la\ muestra\ que\ cumplen\ la\ condición\ C}{n}$
    - Se considera que el estimador estandarizado tiene distribución Normal estándar
  - Tamaño de la muestra = $n$
  - Nivel de error o significación = $\alpha$
  - $z_{\alpha/2} = q\left(1 - \frac{\alpha}{2}\right)$ → Se obtiene con la tabla A4 o con software
    - Aplicado a una variable Normal estándar N(0,1)

# 8. Intervalo de confianza para una proporción poblacional. Fórmulas

| Se conoce la varianza poblacional ($\sigma^2$) | Tamaño de la muestra (n) | Intervalo de confianza |
| --- | --- | --- |
| --- | Grande | $L_i = \hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$<br>$L_s = \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ |

# 8. Intervalo de confianza para una proporción poblacional. Ejemplo

- $p =$ proporción de personas de una población que saben conducir
- Muestra: $n = 100$
- $\alpha = 0.05 \ \rightarrow$ Nivel de confianza del 95%
- Personas de la muestra que saben conducir = 60
  - $\hat{p} = \frac{60}{100} = 0.6$
  - $z_{\alpha/2} = q\left(1 - \frac{\alpha}{2}\right) = q(0.975) = 1.96$
    - Usando un software o la tabla A4
- Intervalo de confianza
  - $L_i = \hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.6 - (1.96)\sqrt{\frac{0.6(1-0.6)}{100}} = 0.5$
  - $L_s = \hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.6 + (1.96)\sqrt{\frac{0.6(1-0.6)}{100}} = 0.7$

# 8. Intervalo de confianza para una proporción poblacional. Ejercicios propuestos

- Ejercicio 9.10(a) del libro
  - La respuesta está disponible en el libro

# 9. Intervalo de confianza para la varianza poblacional

- Parámetro $\theta$ = Varianza poblacional = $\sigma^2$
- Estimador $\hat{\theta}$ = Varianza muestral = $S^2$
  - $\hat{\theta} = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
- La distribución de probabilidad del estimador no es simétrica, por lo que
  - $[L_i, L_s] = \left[\hat{\theta} - \varepsilon_1, \hat{\theta} + \varepsilon_2\right]$ pero $\varepsilon_1 \neq \varepsilon_2$
- Se considera que se ajusta a una distribución Chi-cuadrado con n-1 grados de libertad y valores críticos
  - $\chi^2_{1-\frac{\alpha}{2},n-1}$ y $\chi^2_{\frac{\alpha}{2},n-1}$

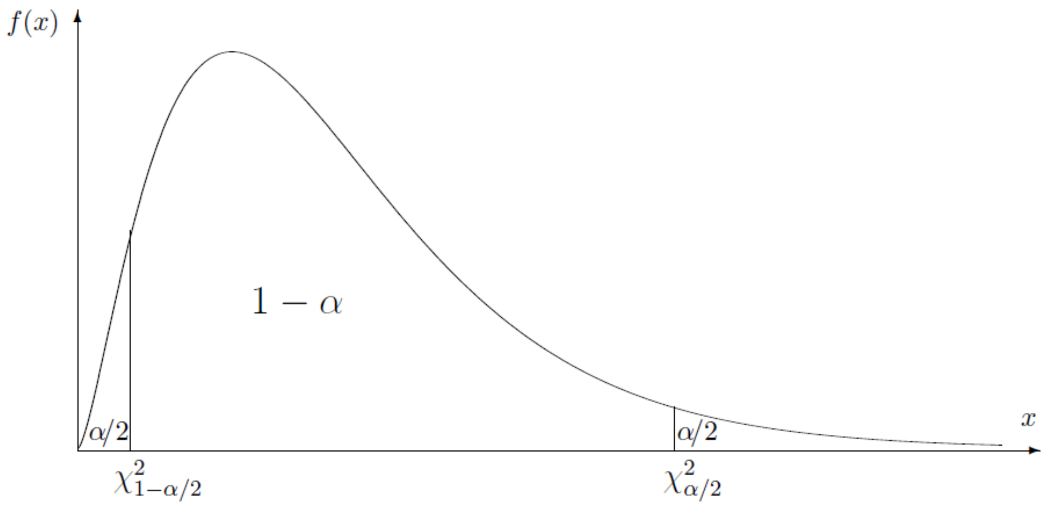


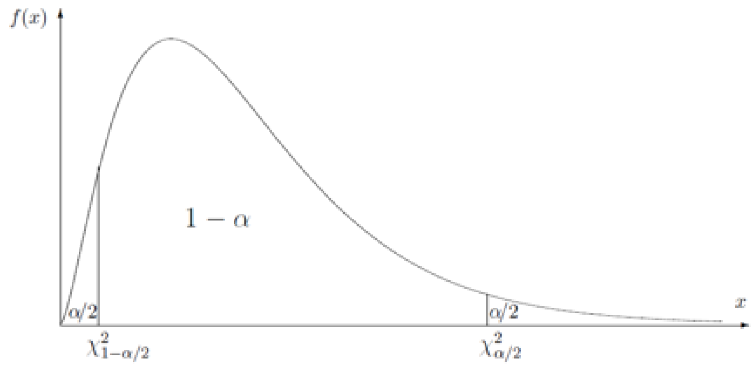
FIGURE 9.13: *Critical values of the Chi-square distribution.*

# 9. Intervalo de confianza para la varianza poblacional. Fórmulas

| Se conoce la varianza poblacional ($\sigma^2$) | Tamaño de la muestra (n) | Intervalo de confianza |
|---|---|---|
| No | Cualquiera | $Li = \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}$ <br> $Ls = \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}$ |

# 9. Intervalo de confianza para la varianza poblacional. Cálculo

- El estimador $\hat{\theta}$ se convierte en una variable Chi-cuadrado $\chi^2(n-1)$
  - $\chi^2(n-1) = \frac{(n-1)S^2}{\sigma^2}$
- Los valores de la variable $\chi^2$ correspondientes a los cuantiles $\frac{\alpha}{2}$ y $1-\frac{\alpha}{2}$ se denominan valores críticos, y se denotan como
  - ${\chi^2}_{1-\frac{\alpha}{2},n-1} = q\left(\frac{\alpha}{2}\right) \rightarrow P\left\{\chi^2 < {\chi^2}_{1-\frac{\alpha}{2},n-1}\right\} = \frac{\alpha}{2}$
  - ${\chi^2}_{\frac{\alpha}{2},n-1} = q\left(1-\frac{\alpha}{2}\right) \rightarrow P\left\{\chi^2 < {\chi^2}_{\frac{\alpha}{2},n-1}\right\} = 1-\frac{\alpha}{2}$
- Por lo que
  - $P\left\{{\chi^2}_{1-\frac{\alpha}{2},n-1} \leq \frac{(n-1)S^2}{\sigma^2} \leq {\chi^2}_{\frac{\alpha}{2},n-1}\right\} = (1-\alpha) \rightarrow P\left\{\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right\} = (1-\alpha)$




- Y el intervalo de confianza usando una muestra concreta con varianza muestral $s^2$ es
  - $\left[\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right]$

# 9. Intervalo de confianza para la varianza poblacional. Tabla Chi-cuadrado

- Para obtener $\chi^2_{1-\frac{\alpha}{2},n-1}$ y $\chi^2_{\frac{\alpha}{2},n-1}$ se puede usar la tabla A6
- Ejemplo
  - $\alpha/2 = 0.05$
  - $1-\alpha/2 = 0.95$
  - Grados de libertad: $n-1 = 5$
- Cálculo de $\chi^2_{0.05,\,5}$
  - Hay que localizar la fila 5
  - Hay que localizar la columna 0.05
  - En la celda está el valor buscado: $\chi^2_{0.05,5} = 11.1$
- Cálculo de $\chi^2_{0.95,\,5}$
  - Hay que localizar la fila 5
  - Hay que localizar la columna 0.95
  - En la celda está el valor buscado: $\chi^2_{0.95,5} = 1.15$

**Table A6. Table of Chi-Square Distribution**

$\chi^2_\alpha$; critical values, such that $P\{\chi^2 > \chi^2_\alpha\} = \alpha$

| $\nu$ | $\alpha$, the right-tail probability | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| (d.f.) | .999 | .995 | .99 | .975 | .95 | .90 | .80 | .20 | .10 | .05 | .025 |
| 1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.06 | 1.64 | 2.71 | 3.84 | 5.02 |
| 2 | 0.00 | 0.01 | 0.02 | 0.05 | 0.10 | 0.21 | 0.45 | 3.22 | 4.61 | 5.99 | 7.38 |
| 3 | 0.02 | 0.07 | 0.11 | 0.22 | 0.35 | 0.58 | 1.01 | 4.64 | 6.25 | 7.81 | 9.35 |
| 4 | 0.09 | 0.21 | 0.30 | 0.48 | 0.71 | 1.06 | 1.65 | 5.99 | 7.78 | 9.49 | 11.1 |
| 5 | 0.21 | 0.41 | 0.55 | 0.83 | 1.15 | 1.61 | 2.34 | 7.29 | 9.24 | 11.1 | 12.8 |

# 9. Intervalo de confianza para la varianza poblacional. Ejemplo 9.40

- Construir un intervalo de confianza del 90 % para la varianza poblacional basado en una muestra de 6 mediciones
  - 2.5, 7.4, 8.0, 4.5, 7.4, 9.2

# 9. Intervalo de confianza para la varianza poblacional. Ejemplo 9.40 (Solución)

- Estimador $\hat{\theta}$ = Varianza muestral = $S^2$
  - $n = 6$
  - $\bar{x} = \frac{2.5+7.4+8.0+4.5+7.4+9.2}{6} = 6.5$
  - $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{6}\left((2.5-6.5)^2 + \cdots + (9.2-6.5)^2\right) = 6.23$
  - $\alpha = 1 - 0.90 = 0.1$
  - ${\chi^2}_{1-\frac{\alpha}{2},n-1} = {\chi^2}_{0.95,\,5} = 1.15$ (Se calcular usando un software como R o la tabla A6 de libro)
  - ${\chi^2}_{\frac{\alpha}{2},n-1} = {\chi^2}_{0.05,\,5} = 11.1$
- Intervalo de confianza para la varianza poblacional
  - $Li = \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}} = \frac{(6-1)(6.23)}{11.1} = 2.8$
  - $Ls = \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}} = \frac{(6-1)(6.23)}{1.15} = 27.1$
- Intervalo para la desviación estándar poblacional
  - $\left[\sqrt{2.8}, \sqrt{27.1}\right] = [1.68, 5.21]$

# 9. Intervalo de confianza para la varianza poblacional. Ejercicios propuestos

- Ejercicio 9.23(c) del libro

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones

- Si existen dos poblaciones, se puede construir un intervalo de confianza para la diferencia entre las medias poblacionales de ambas
  - Cuando se dispone de muestras grandes
  - O cuando se dispone de muestras de pequeño tamaño con distribución Normal

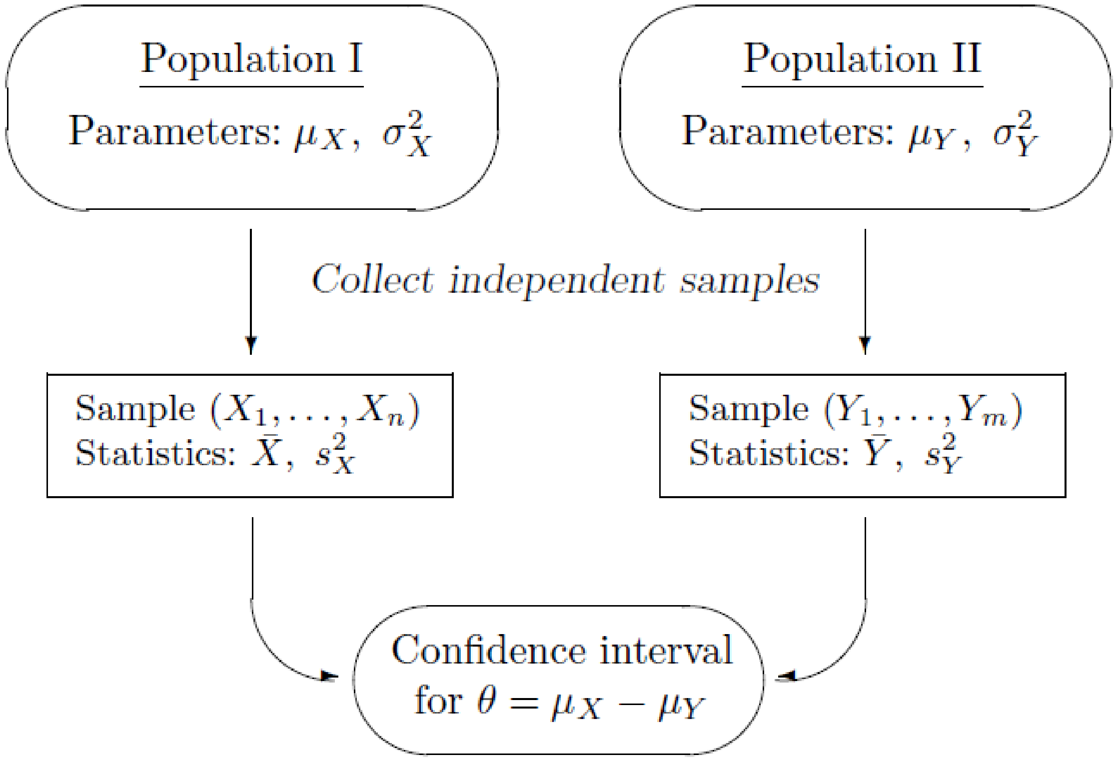



FIGURE 9.4: *Comparison of two populations.*

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones

- Casos posibles
  - Varianzas poblacionales conocidas → Se utiliza un estimador con distribución Normal
  - Varianzas poblacionales desconocidas
    - Muestras grandes → Se utiliza un estimador con distribución Normal
    - Muestras pequeñas con distribución Normal
      - Varianzas iguales → Se utiliza un estimador con distribución T de Student
      - Varianzas diferentes → Se utiliza un estimador con distribución aproximada a T de Student
    - Muestras pequeñas sin distribución Normal o desconocida → Se aplican técnicas de inferencia no paramétrica (ver capítulo 10.2 del libro)
- Nota
  - Una muestra se suele considerar grande cuando el tamaño es $n \geq 30$ o $n > 30$ (no hay unanimidad).

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones. Interpretación

- Si el intervalo de confianza para la diferencia de las medias de dos poblaciones contiene el valor 0, se puede para afirmar que la diferencia de las medias NO es significativa.
- Si el intervalo no contiene el valor 0, se puede afirmar que la diferencia entre las medias de ambas poblaciones SÍ es significativa y además:
  - Si los extremos del intervalo son negativos, se puede afirmar que el parámetro de la primera población es menor que el de la segunda.
  - Si los extremos del intervalo son positivos, se puede afirmar que el parámetro de la primera población es mayor que el de la segunda.

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones. Fórmulas (1)

| **Se conocen las varianzas poblacionales ($\sigma_X^2, \sigma_Y^2$)** | **Tamaño de las muestras (n, m)** | **Intervalo de confianza** |
|---|---|---|
| Si | Cualquiera | $L_i = \bar{x} - \bar{y} - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$ <br> $L_s = \bar{x} - \bar{y} + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$ |
| No | Grande | Se sustituyen las varianzas poblacionales por las muestrales <br> $L_i = \bar{x} - \bar{y} - z_{\frac{\alpha}{2}}\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}$ <br> $L_s = \bar{x} - \bar{y} + z_{\frac{\alpha}{2}}\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}$ |

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones. Fórmulas (2)

| Se conocen las varianzas poblacionales ($\sigma_X^2, \sigma_Y^2$) | Tamaño de las muestras (n, m) | Intervalo de confianza |
|---|---|---|
| No<br>Pero son iguales | Muestras pequeñas con distribución Normal | $L_i = \bar{x} - \bar{y} - (t_{\frac{\alpha}{2}, n+m-2})(s_p)\sqrt{\frac{1}{n} + \frac{1}{m}}$<br>$L_s = \bar{x} - \bar{y} + (t_{\frac{\alpha}{2}, n+m-2})(s_p)\sqrt{\frac{1}{n} + \frac{1}{m}}$<br><br>$n + m - 2$ = grados de libertad<br>$s_p$= varianza muestral agrupada<br><br>$s_p = \sqrt{\frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}}$ |

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones. Fórmulas (3)

| **Se conocen las varianzas poblacionales ($\sigma_X^2, \sigma_Y^2$)** | **Tamaño de las muestras (n, m)** | **Intervalo de confianza** |
|---|---|---|
| No<br>Pero son diferentes | Muestras pequeñas con distribución Normal | $L_i = \bar{x} - \bar{y} - t_{\frac{\alpha}{2},\nu}\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}$<br>$L_s = \bar{x} - \bar{y} + t_{\frac{\alpha}{2},\nu}\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}$<br><br>$\nu$ = grados de libertad<br>$\nu = \frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)^2}{\frac{s_X^4}{n^2(n-1)} + \frac{s_Y^4}{m^2(m-1)}}$ |

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones. Ejemplo 9.14

- Un experto evalúa la efectividad de una actualización del hardware de un ordenador ejecutando un cierto programa 50 veces antes de la actualización y 50 veces después de ella.
- Según estos datos, el tiempo de ejecución promedio es de
  - 8.5 minutos antes de la actualización
  - 7.22 minutos después de la actualización
- Históricamente, la desviación estándar ha sido de 1.8 minutos, y presumiblemente no ha cambiado.
- Construir un intervalo de confianza del 90 % que muestre cuánto se redujo el tiempo de ejecución medio debido a la actualización del hardware.

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones. Ejemplo 9.14 (Solución)

- Es un caso de muestras grandes con varianzas conocidas
- X = tiempo de ejecución en una muestra de 50 ejecuciones antes de la actualización
- Y = tiempo de ejecución en una muestra de 50 ejecuciones después de la actualización
- $n = m = 50$
- $\sigma_X = \sigma_Y = 1.8 \text{ min}$
- $\bar{x} = 8.5 \text{ min}$
- $\bar{y} = 7.2 \text{ min}$
- $1 - \alpha = 0.9 \rightarrow \alpha = 0.1 \rightarrow \frac{\alpha}{2} = 0.05$
- Según la tabla A4: $z_{\frac{\alpha}{2}} = 1.64$
- Intervalo de confianza→ Como no contiene el valor 0, la diferencia de medias SÍ es significativa, siendo mayor la media de la primera población, por ser los extremos positivos
  - $L_i = \bar{x} - \bar{y} - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} = 8.5 - 7.2 - 1.64\sqrt{\frac{1.8^2}{50} + \frac{1.8^2}{50}} = 0.7$
  - $L_s = \bar{x} - \bar{y} + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} = 8.5 + 7.2 - 1.64\sqrt{\frac{1.8^2}{50} + \frac{1.8^2}{50}} = 1.9$

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones. Ejemplo 9.16

- Las conexiones a Internet a menudo se ralentizan por retrasos en los nodos
- Hay que comprobar si el tiempo de retardo aumenta durante los tiempos de volumen pesado
- Quinientos paquetes se envían a través de la misma red entre las 5 pm y las 6 pm
- Trescientos paquetes se envían entre las 10 pm y las 11 pm
- La muestra temprana tiene un tiempo de retardo medio de 0,8 segundos con una desviación estándar de 0,1 seg
- La segunda muestra tiene un tiempo de retardo medio de 0,5 seg con una desviación estándar de 0,08 seg
- Construir un intervalo de confianza del 99,5 % para la diferencia entre los tiempos de retardo medios

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones. Ejemplo 9.16 (Solución)

- Es un caso de muestras grandes con varianzas desconocidas
- X = tiempo de retado en una muestra de 500 paquetes que se envían entre las 5 pm y las 6 pm
- Y = tiempo de retardo en una muestra de 300 paquetes se envían entre las 10 pm y las 11 pm
- $n = 500$ m$= 300$
- $\bar{x} = 0.8 \text{ seg}$
- $\bar{y} = 0.5 \text{ seg}$
- $s_X = 0.1 \text{ seg}$
- $s_Y = 0.08 \text{ seg}$
- $1 - \alpha = 0.995 \rightarrow \alpha = 0.005 \rightarrow \frac{\alpha}{2} = 0.0025$
- Según la tabla A4: $z_{\frac{\alpha}{2}} = 2.81$
- Intervalo de confianza→ Como no contiene el valor 0, la diferencia de medias SÍ es significativa, siendo mayor la media de la primera población, por ser los extremos positivos

  - $L_i = \bar{x} - \bar{y} - z_{\frac{\alpha}{2}}\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}} = 0.8 - 0.5 - 2.81\sqrt{\frac{0.1^2}{500} + \frac{0.08^2}{300}} = 0.28 \text{ seg}$

  - $L_s = \bar{x} - \bar{y} + z_{\frac{\alpha}{2}}\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}} = 0.8 - 0.5 + 2.81\sqrt{\frac{0.1^2}{500} + \frac{0.08^2}{300}} = 0.32 \text{ seg}$

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones. Ejemplo 9.20

- La grabación de CD (Compact Disc) consume energía y afecta a la vida útil de la batería en los ordenadores portátiles.
- Para estimar el efecto de la escritura en CD, se pide a 30 usuarios que trabajen en sus ordenadores portátiles hasta que aparezca el signo de "baja batería".
- Dieciocho usuarios sin CD trabajaron un promedio de 5,3 horas con una desviación estándar de 1,4 horas.
- Los otros doce, que utilizaron su grabador de CD, trabajaron un promedio de 4,8 horas con una desviación estándar de 1,6 horas.
- Suponiendo distribuciones Normales con las mismas varianzas poblacionales, construir un intervalo de confianza del 95 % para la reducción de la vida útil de la batería causada por la grabación de CD.

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones. Ejemplo 9.20 (Solución)

- Es un caso de muestras pequeñas (<30) con varianzas desconocidas pero iguales
- X = tiempo de uso en una muestra de 12 usuarios con CD
- Y = tiempo de uso en una muestra de 18 usuarios sin CD
- $n = 12$ m$= 18$
- $\bar{x} = 4.8$ horas
- $\bar{y} = 5.3$ horas
- $s_X = 1.6$ horas
- $s_Y = 1.4$ horas
- $1 - \alpha = 0.95 \rightarrow \alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025$
- Según la tabla A5: $t_{\frac{\alpha}{2}, n+m-2} = t_{0.025,\,28} = 2.05$
- $s_p = \sqrt{\frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}} = \sqrt{\frac{(11)(1.6)^2 + (17)(1.4)^2}{12+18-2}} = 1.48$
- Intervalo de confianza → Como contiene el valor 0, la diferencia de medias NO es significativa
  - $L_i = \bar{x} - \bar{y} - \left(t_{\frac{\alpha}{2}, n+m-2}\right)\left(s_p\right)\sqrt{\frac{1}{n} + \frac{1}{m}} = 4.8 - 5.3 - (2.05)(1.48)\sqrt{\frac{1}{12} + \frac{1}{18}} = -1.63$ horas
  - $L_s = \bar{x} - \bar{y} - \left(t_{\frac{\alpha}{2}, n+m-2}\right)\left(s_p\right)\sqrt{\frac{1}{n} + \frac{1}{m}} = 4.8 + 5.3 - (2.05)(1.48)\sqrt{\frac{1}{12} + \frac{1}{18}} = 0.63$ horas

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones. Ejemplo 9.21

- Una cuenta en el servidor A es más cara que una cuenta en el servidor B.
- Sin embargo, el servidor A es más rápido.
- Para ver si es óptimo comprar el servidor más rápido, pero más caro, un administrador necesita saber cuánto más rápido es.
- Un determinado algoritmo informático se ejecuta 30 veces en el servidor A y 20 veces en el servidor B con los siguientes resultados:

| | **Servidor A** | **Servidor B** |
|---|---|---|
| **Media muestral** | 6.7 min | 7.5 min |
| **Deviación estándar muestral** | 0.6 min | 1.2 min |

- Construir un intervalo de confianza del 95 % para la diferencia entre los tiempos de ejecución promedio en el servidor A y el servidor B, asumiendo que los tiempos observados son aproximadamente normales.

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones. Ejemplo 9.21 (Solución)

- Es un caso de muestras pequeñas (porque hay una <30) pero con distribución Normal, con varianzas desconocidas y diferentes
- X = tiempo de ejecución del algoritmo en una muestra de 30 ejecuciones
- Y = tiempo de ejecución del algoritmo en una muestra de 20 ejecuciones
- $n = 30$ m$= 20$
- $\bar{x} = 6.7 \text{ min}$
- $\bar{y} = 7.5 \text{ min}$
- $s_X = 0.6 \text{ min}$
- $s_Y = 1.2 \text{ min}$
- $1 - \alpha = 0.95 \rightarrow \alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025$
- $\nu = \frac{\left(\frac{s_X^2}{n}+\frac{s_Y^2}{m}\right)^2}{\frac{s_X^4}{n^2(n-1)}+\frac{s_Y^4}{m^2(m-1)}} = \frac{\left(\frac{0.6^2}{30}+\frac{1.2^2}{20}\right)^2}{\frac{0.6^4}{30^2(30-1)}+\frac{1.2^4}{20^2(20-1)}} = 25.4 \approx 25$
- Según la tabla A5: $t_{\frac{\alpha}{2},\nu} = t_{0.025,\,25} = 2.06$
- Intervalo de confianza → Como no contiene el valor 0, la diferencia de medias SÍ es significativa, siendo mayor la de la segunda población
  - $L_i = \bar{x} - \bar{y} - t_{\frac{\alpha}{2},\nu}\sqrt{\frac{s_X^2}{n}+\frac{s_Y^2}{m}} = 6.7 - 7.5 - 2.06\sqrt{\frac{0.6^2}{30}+\frac{1.2^2}{20}} = -1.4 \text{ min}$
  - $L_s = \bar{x} - \bar{y} - t_{\frac{\alpha}{2},\nu}\sqrt{\frac{s_X^2}{n}+\frac{s_Y^2}{m}} = 6.7 - 7.5 + 2.06\sqrt{\frac{0.6^2}{30}+\frac{1.2^2}{20}} = -0.24 \text{ min}$

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones. Muestras pareadas (1)

- Si las muestras no son independientes, se dice que son dependientes o pareadas
- En las dos muestras se tienen los valores de una propiedad de la población para los mismos elementos de la población, en dos situaciones diferentes
- Ejemplo
  - Una muestra puede ser el peso de unas personas antes de una dieta, y la otra muestra el peso de las mismas personas antes de la dieta
- En este caso se restan uno a uno los valores de las dos muestras correspondientes al mismo elemento, y se aplican las fórmulas del intervalo de confianza para la media de una población sobre la variable aleatoria (D) que se crea como diferencia entre los valores de las dos muestras
  - $X = \{X_1, X_2, \ldots, X_n\}$
  - $Y = \{Y_1, Y_2, \ldots, Y_n\}$
  - $D = \{(X_1 - Y_1), (X_2 - Y_2), \ldots, (X_n - Y_n)\}$

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones. Muestras pareadas (2)

| **Tamaño de las muestras (n)** | **Intervalo de confianza** |
|---|---|
| Grande | $L_i = \bar{d} - z_{\frac{\alpha}{2}} \frac{s_D}{\sqrt{n}}$ <br> $L_s = \bar{d} + z_{\frac{\alpha}{2}} \frac{s_D}{\sqrt{n}}$ |
| Pequeña con distribución Normal | $L_i = \bar{d} - t_{\frac{\alpha}{2},n-1} \frac{s_D}{\sqrt{n}}$ <br> $L_s = \bar{d} + t_{\frac{\alpha}{2},n-1} \frac{s_D}{\sqrt{n}}$ |

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones. Muestras pareadas. Ejemplo

- Ocho personas realizan una dieta, los pesos antes y después de la dieta son las siguientes muestras, en el mismo orden de personas en cada una
  - $x = (65.4, 68.0, 68.5, 68.6, 69.0, 69.1, 71.0, 71.5)$
  - $y = (65.0, 65.4, 66.2, 66.7, 68.1, 70.3, 72.8, 75.4)$
- Construir un intervalo de confianza, con un nivel de confianza del 95%

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones. Muestras pareadas. Ejemplo (Solución)

- Es un caso de muestras pareadas pequeñas, pero con distribución Normal
- X = peso de una persona antes de una dieta
- Y = peso de una persona después de la dieta
- $d = \{(65.4 - 65.0), (68.0 - 65.4), \dots, (71.5 - 75.4)\} = \{0.4,\ 2.6,\ 2.3,\ 1.9, 0.9, -1.2, -1.8, -3.9\}$
- $s_D = 2.28\ \mathrm{kg}$
- $\bar{d} = 0.15\ \mathrm{kg}$
- $1 - \alpha = 0.95 \rightarrow \alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025$
- Según la tabla A5: $t_{\frac{\alpha}{2}, n-1} = t_{0.025,\, 7} = 2.36$
- Intervalo de confianza → Como contiene el valor 0, la diferencia de medias NO es significativa
  - $L_i = \bar{d} - t_{\frac{\alpha}{2}, n-1} \frac{s_D}{\sqrt{n}} = 0.15 - 2.36 \frac{2.28}{\sqrt{8}} = -1.75\ \mathrm{kg}$
  - $L_s = \bar{d} + t_{\frac{\alpha}{2}, n-1} \frac{s_D}{\sqrt{n}} = 0.15 + 2.36 \frac{2.28}{\sqrt{8}} = 2.05\ \mathrm{kg}$

# 10. Intervalo de confianza para la diferencia de las medias de dos poblaciones. Ejercicios propuestos

- Ejercicios 9.18(a) y 9.24(b) del libro
  - La respuesta de 9.18 está disponible en el libro

# 11. Intervalo de confianza para la diferencia de proporciones en dos poblaciones

- Parámetros
  - Proporción poblacional en la primera población: $p_1$
  - Proporción poblacional en la segunda población: $p_2$
- Muestras
  - Tamaño de la muestra de la primera población: $n_1$
  - Tamaño de la muestra de la segunda población: $n_2$
- Estimadores
  - Proporción muestral de la primera población: $\hat{P}_1$
  - Proporción muestral de la segunda población: $\hat{P}_2$

# 11. Intervalo de confianza para la diferencia de proporciones en dos poblaciones. Fórmulas

| Tamaño de las muestras (n1,n2) | Intervalo de confianza |
| --- | --- |
| Grande | $L_i = \hat{p}_1 - \hat{p}_2 - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ <br> $L_s = \hat{p}_1 - \hat{p}_2 - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ |

# 11. Intervalo de confianza para la diferencia de proporciones en dos poblaciones. Ejemplo 9.17

- Un candidato se prepara para las elecciones locales.
- En su campaña, 42 de 70 personas seleccionadas al azar en la ciudad A y 59 de las 100 personas seleccionadas al azar en la ciudad B mostraron que votarían por este candidato.
- Estimar la diferencia de apoyo que este candidato está recibiendo en las ciudades A y B con un 95 % de confianza.

# 11. Intervalo de confianza para la diferencia de proporciones en dos poblaciones. Ejemplo 9.17 (Solución)

- Son muestras grandes: $n_1 = 70, \quad n_2 = 100$
- $\hat{p}_1 = \frac{42}{70} = 0.6$
- $\hat{p}_2 = \frac{59}{100} = 0.59$
- $1 - \alpha = 0.95 \rightarrow \alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025$
- Según la tabla A4: $z_{\frac{\alpha}{2}} = z_{0.025} = q(1 - 0.025) = q(0.975) = 1.96$
- Intervalo de confianza → Como contiene el valor 0, la diferencia NO es significativa
  - $L_i = \hat{p}_1 - \hat{p}_2 - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = 0.6 - 0.59 - 1.96\sqrt{\frac{0.6(1-0.6)}{70} + \frac{0.59(1-0.59)}{100}} = -0.14$
  - $L_s = \hat{p}_1 - \hat{p}_2 + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = 0.6 - 0.59 + 1.96\sqrt{\frac{0.6(1-0.6)}{70} + \frac{0.59(1-0.59)}{100}} = 0.16$

# 11. Intervalo de confianza para la diferencia de proporciones. Ejercicios propuestos

- Ejercicio 9.16(a) y 9.17 del libro
  - La respuesta de 9.17 está disponible en el libro

# 12. Resumen

- La inferencia estadística permite encontrar proponer valores para los parámetros de una población, cuando no se pueden calcular directamente
- El resultado de la inferencia estadística puede ser un intervalo de confianza, la comprobación de una hipótesis, o un modelo de predicción
- La inferencia estadística se basa en el uso de estimadores para inferir el valor de los parámetros de una población
- Los estimadores se calculan sobre una muestra de la población
- Un parámetro no es una variable aleatoria, tiene un valor único desconocido, pero un estimador sí es una variable aleatoria, porque su valor puede ser diferente, en diferentes muestras de la misma población
- La estimación de un parámetro puede ser puntual o por intervalo
- Los estimadores más usados son los estadísticos de la muestra, especialmente la media muestral, la varianza muestral y la proporción muestral
- Un intervalo de confianza para un parámetro de una población contiene el valor del parámetro con una probabilidad igual a un nivel de confianza preestablecido
- Se pueden construir intervalos de confianza para los parámetros de una población, como la media, la varianza o las proporciones
- Y también se pueden construir intervalos de confianza para la diferencia entre dos mismos parámetros de dos poblaciones diferentes
- En el cálculo de intervalos de confianza se aplican distribuciones de probabilidad Normal, T de Student y Chi-cuadrado