

Probabilidad y Estadística para Inteligencia Artificial

Dr. Ing. Pablo Briff

Laboratorio de Sistemas Embebidos - FIUBA

pbriff@fi.uba.ar

1 de Agosto de 2020



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Tabla de Contenidos I

1 Estimación de Intervalo

- Estimación de Intervalo
- Distribución Chi-Cuadrado
- Distribución t de Student
- Estimación de Intervalo con Varianza Desconocida
- Estimación de Intervalo con Varianza Desconocida
- Ejemplo de Intervalo de Confianza
- Test de Hipótesis

2 Ejercicios Práctico-Teóricos

- Ejercicio 1
- Ejercicio 2
- Ejercicio 3
- Ejercicio 4

3 Bibliografía



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

- La estimación puntual estima un parámetro θ



Estimación de Intervalo

- La estimación puntual estima un parámetro θ
- En la estimación por intervalo queremos estimar el intervalo en el cual está θ con un cierto grado de confianza



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo

- La estimación puntual estima un parámetro θ
- En la estimación por intervalo queremos estimar el intervalo en el cual está θ con un cierto grado de confianza
- A esto se lo denomina Intervalo de Confianza



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo

- La estimación puntual estima un parámetro θ
- En la estimación por intervalo queremos estimar el intervalo en el cual está θ con un cierto grado de confianza
- A esto se lo denomina Intervalo de Confianza
- Llamaremos α al nivel de significación



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo

- La estimación puntual estima un parámetro θ
- En la estimación por intervalo queremos estimar el intervalo en el cual está θ con un cierto grado de confianza
- A esto se lo denomina Intervalo de Confianza
- Llamaremos α al nivel de significación
- Llamaremos a $1 - \alpha$ el nivel de confianza



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo

- La estimación puntual estima un parámetro θ
- En la estimación por intervalo queremos estimar el intervalo en el cual está θ con un cierto grado de confianza
- A esto se lo denomina Intervalo de Confianza
- Llamaremos α al nivel de significación
- Llamaremos a $1 - \alpha$ el nivel de confianza
- Para estimar el intervalo, vamos a usar la pdf del estimador puntual



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

- Sea $X = \{x_i\}_{i=1}^N$ una muestra de N puntos de una v.a. normal con media μ (desconocida) y varianza σ^2 (conocida)



Estimación de Intervalo

- Sea $X = \{x_i\}_{i=1}^N$ una muestra de N puntos de una v.a. normal con media μ (desconocida) y varianza σ^2 (conocida)
- Queremos estimar la media μ con la media muestral $\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo

- Sea $X = \{x_i\}_{i=1}^N$ una muestra de N puntos de una v.a. normal con media μ (desconocida) y varianza σ^2 (conocida)
- Queremos estimar la media μ con la media muestral $\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$
- $\hat{\mu} \sim \mathcal{N}(\mu, \frac{\sigma^2}{N})$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo

- Sea $X = \{x_i\}_{i=1}^N$ una muestra de N puntos de una v.a. normal con media μ (desconocida) y varianza σ^2 (conocida)
- Queremos estimar la media μ con la media muestral $\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$
- $\hat{\mu} \sim \mathcal{N}(\mu, \frac{\sigma^2}{N})$
- Definimos la siguiente estadística:



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo

- Sea $X = \{x_i\}_{i=1}^N$ una muestra de N puntos de una v.a. normal con media μ (desconocida) y varianza σ^2 (conocida)
- Queremos estimar la media μ con la media muestral $\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$
- $\hat{\mu} \sim \mathcal{N}(\mu, \frac{\sigma^2}{N})$
- Definimos la siguiente estadística:
- $Z = \frac{\hat{\mu} - \mu}{\sigma / \sqrt{N}} \sim \mathcal{N}(0, 1)$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo

- Sea $X = \{x_i\}_{i=1}^N$ una muestra de N puntos de una v.a. normal con media μ (desconocida) y varianza σ^2 (conocida)
- Queremos estimar la media μ con la media muestral $\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$
- $\hat{\mu} \sim \mathcal{N}(\mu, \frac{\sigma^2}{N})$
- Definimos la siguiente estadística:
- $Z = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{N}} \sim \mathcal{N}(0, 1)$
- $P(-1.96 < Z < 1.96) = 0.95$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo

- Sea $X = \{x_i\}_{i=1}^N$ una muestra de N puntos de una v.a. normal con media μ (desconocida) y varianza σ^2 (conocida)
- Queremos estimar la media μ con la media muestral $\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$
- $\hat{\mu} \sim \mathcal{N}(\mu, \frac{\sigma^2}{N})$
- Definimos la siguiente estadística:
- $Z = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{N}} \sim \mathcal{N}(0, 1)$
- $P(-1.96 < Z < 1.96) = 0.95$
- En otras palabras, Z está en ese intervalo con una confianza del 95%



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

- Reemplazando queda:



- Reemplazando queda:
- $P(-1.96 < \sqrt{N} \frac{\hat{\mu} - \mu}{\sigma} < 1.96) = 0.95$



- Reemplazando queda:
- $P(-1.96 < \sqrt{N} \frac{\hat{\mu} - \mu}{\sigma} < 1.96) = 0.95$
- Despejando μ :



Estimación de Intervalo

- Reemplazando queda:
- $P(-1.96 < \sqrt{N} \frac{\hat{\mu} - \mu}{\sigma} < 1.96) = 0.95$
- Despejando μ :
- $P(\hat{\mu} - 1.96 \frac{\sigma}{\sqrt{N}} < \mu < \hat{\mu} + 1.96 \frac{\sigma}{\sqrt{N}}) = 0.95$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo

- Reemplazando queda:
- $P(-1.96 < \sqrt{N} \frac{\hat{\mu} - \mu}{\sigma} < 1.96) = 0.95$
- Despejando μ :
- $P(\hat{\mu} - 1.96 \frac{\sigma}{\sqrt{N}} < \mu < \hat{\mu} + 1.96 \frac{\sigma}{\sqrt{N}}) = 0.95$
- Este es un intervalo de confianza de dos lados (colas de la normal por exceso y por defecto)



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo

- Reemplazando queda:
- $P(-1.96 < \sqrt{N} \frac{\hat{\mu} - \mu}{\sigma} < 1.96) = 0.95$
- Despejando μ :
- $P(\hat{\mu} - 1.96 \frac{\sigma}{\sqrt{N}} < \mu < \hat{\mu} + 1.96 \frac{\sigma}{\sqrt{N}}) = 0.95$
- Este es un intervalo de confianza de dos lados (colas de la normal por exceso y por defecto)
- Se puede calcular el intervalo de confianza para un valor arbitrario dado



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo

- Reemplazando queda:
- $P(-1.96 < \sqrt{N} \frac{\hat{\mu} - \mu}{\sigma} < 1.96) = 0.95$
- Despejando μ :
- $P(\hat{\mu} - 1.96 \frac{\sigma}{\sqrt{N}} < \mu < \hat{\mu} + 1.96 \frac{\sigma}{\sqrt{N}}) = 0.95$
- Este es un intervalo de confianza de dos lados (colas de la normal por exceso y por defecto)
- Se puede calcular el intervalo de confianza para un valor arbitrario dado
- A medida que aumenta la confianza, el intervalo se hace más grande



Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$
- Dado que Z es simétrica alrededor de la media, se cumple



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$
- Dado que Z es simétrica alrededor de la media, se cumple
- $z_{1-\alpha/2} = -z_{\alpha/2}$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$
- Dado que Z es simétrica alrededor de la media, se cumple
- $z_{1-\alpha/2} = -z_{\alpha/2}$
- $P(Z < -z_{\alpha/2}) = P(Z > z_{\alpha/2}) = \alpha/2$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$
- Dado que Z es simétrica alrededor de la media, se cumple
- $z_{1-\alpha/2} = -z_{\alpha/2}$
- $P(Z < -z_{\alpha/2}) = P(Z > z_{\alpha/2}) = \alpha/2$
- Para $\alpha = 0.05$, $P(Z < -z_{\alpha/2}) = \text{normcdf}(-1.96) = 0.025 = \alpha/2$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$
- Dado que Z es simétrica alrededor de la media, se cumple
- $z_{1-\alpha/2} = -z_{\alpha/2}$
- $P(Z < -z_{\alpha/2}) = P(Z > z_{\alpha/2}) = \alpha/2$
- Para $\alpha = 0.05$, $P(Z < -z_{\alpha/2}) = \text{normcdf}(-1.96) = 0.025 = \alpha/2$
- $P(Z > z_{\alpha/2}) = \text{normcdf}(1.96) = 0.975 = 1 - \alpha/2$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$
- Dado que Z es simétrica alrededor de la media, se cumple
- $z_{1-\alpha/2} = -z_{\alpha/2}$
- $P(Z < -z_{\alpha/2}) = P(Z > z_{\alpha/2}) = \alpha/2$
- Para $\alpha = 0.05$, $P(Z < -z_{\alpha/2}) = \text{normcdf}(-1.96) = 0.025 = \alpha/2$
- $P(Z > z_{\alpha/2}) = \text{normcdf}(1.96) = 0.975 = 1 - \alpha/2$
- Para un intervalo de confianza $1 - \alpha$ se cumple:



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$
- Dado que Z es simétrica alrededor de la media, se cumple
- $z_{1-\alpha/2} = -z_{\alpha/2}$
- $P(Z < -z_{\alpha/2}) = P(Z > z_{\alpha/2}) = \alpha/2$
- Para $\alpha = 0.05$, $P(Z < -z_{\alpha/2}) = \text{normcdf}(-1.96) = 0.025 = \alpha/2$
- $P(Z > z_{\alpha/2}) = \text{normcdf}(1.96) = 0.975 = 1 - \alpha/2$
- Para un intervalo de confianza $1 - \alpha$ se cumple:
- $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$
- Dado que Z es simétrica alrededor de la media, se cumple
- $z_{1-\alpha/2} = -z_{\alpha/2}$
- $P(Z < -z_{\alpha/2}) = P(Z > z_{\alpha/2}) = \alpha/2$
- Para $\alpha = 0.05$, $P(Z < -z_{\alpha/2}) = \text{normcdf}(-1.96) = 0.025 = \alpha/2$
- $P(Z > z_{\alpha/2}) = \text{normcdf}(1.96) = 0.975 = 1 - \alpha/2$
- Para un intervalo de confianza $1 - \alpha$ se cumple:
- $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$
- $P(\hat{\mu} - z_{\alpha/2} \frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2} \frac{\sigma}{\sqrt{N}}) = 1 - \alpha$



Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$
- Dado que Z es simétrica alrededor de la media, se cumple
- $z_{1-\alpha/2} = -z_{\alpha/2}$
- $P(Z < -z_{\alpha/2}) = P(Z > z_{\alpha/2}) = \alpha/2$
- Para $\alpha = 0.05$, $P(Z < -z_{\alpha/2}) = \text{normcdf}(-1.96) = 0.025 = \alpha/2$
- $P(Z > z_{\alpha/2}) = \text{normcdf}(1.96) = 0.975 = 1 - \alpha/2$
- Para un intervalo de confianza $1 - \alpha$ se cumple:
- $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$
- $P(\hat{\mu} - z_{\alpha/2} \frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2} \frac{\sigma}{\sqrt{N}}) = 1 - \alpha$
- Entonces al intervalo $\mu \in (\hat{\mu} - 1.96 \frac{\sigma}{\sqrt{N}}, \hat{\mu} + 1.96 \frac{\sigma}{\sqrt{N}})$ se lo llama *intervalo de confianza de 95% bilateral*



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Intervalo de Confianza Unilateral

- $P(\hat{\mu} - z_{\alpha/2} \frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2} \frac{\sigma}{\sqrt{N}}) = 1 - \alpha$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Intervalo de Confianza Unilateral

- $P(\hat{\mu} - z_{\alpha/2} \frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2} \frac{\sigma}{\sqrt{N}}) = 1 - \alpha$
- El intervalo de confianza porcentual $100 \times (1 - \alpha)$ está dentro de $z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$ unidades de la media muestral $\hat{\mu}$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Intervalo de Confianza Unilateral

- $P(\hat{\mu} - z_{\alpha/2} \frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2} \frac{\sigma}{\sqrt{N}}) = 1 - \alpha$
- El intervalo de confianza porcentual $100 \times (1 - \alpha)$ está dentro de $z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$ unidades de la media muestral $\hat{\mu}$
- Dado que $P(Z < 1.64) = 0.95$, tenemos



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Intervalo de Confianza Unilateral

- $P(\hat{\mu} - z_{\alpha/2} \frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2} \frac{\sigma}{\sqrt{N}}) = 1 - \alpha$
- El intervalo de confianza porcentual $100 \times (1 - \alpha)$ está dentro de $z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$ unidades de la media muestral $\hat{\mu}$
- Dado que $P(Z < 1.64) = 0.95$, tenemos
- $P(\hat{\mu} - 1.64 \frac{\sigma}{\sqrt{N}} < \mu) = 0.95$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Intervalo de Confianza Unilateral

- $P(\hat{\mu} - z_{\alpha/2} \frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2} \frac{\sigma}{\sqrt{N}}) = 1 - \alpha$
- El intervalo de confianza porcentual $100 \times (1 - \alpha)$ está dentro de $z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$ unidades de la media muestral $\hat{\mu}$
- Dado que $P(Z < 1.64) = 0.95$, tenemos
- $P(\hat{\mu} - 1.64 \frac{\sigma}{\sqrt{N}} < \mu) = 0.95$
- Entonces al intervalo $(\hat{\mu} - 1.64 \frac{\sigma}{\sqrt{N}}, \infty)$ se lo llama *intervalo de confianza de 95% superior unilateral*



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Intervalo de Confianza Unilateral

- $P(\hat{\mu} - z_{\alpha/2} \frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2} \frac{\sigma}{\sqrt{N}}) = 1 - \alpha$
- El intervalo de confianza porcentual $100 \times (1 - \alpha)$ está dentro de $z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$ unidades de la media muestral $\hat{\mu}$
- Dado que $P(Z < 1.64) = 0.95$, tenemos
- $P(\hat{\mu} - 1.64 \frac{\sigma}{\sqrt{N}} < \mu) = 0.95$
- Entonces al intervalo $(\hat{\mu} - 1.64 \frac{\sigma}{\sqrt{N}}, \infty)$ se lo llama *intervalo de confianza de 95% superior unilateral*
- Podemos definir análogamente el intervalo de confianza de 95% inferior unilateral como $\mu \in (-\infty, \hat{\mu} + 1.64 \frac{\sigma}{\sqrt{N}})$



FACULTAD
DE INGENIERIA
Universidad de Buenos Aires

Intervalo de Confianza Unilateral

- $P(\hat{\mu} - z_{\alpha/2} \frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2} \frac{\sigma}{\sqrt{N}}) = 1 - \alpha$
- El intervalo de confianza porcentual $100 \times (1 - \alpha)$ está dentro de $z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$ unidades de la media muestral $\hat{\mu}$
- Dado que $P(Z < 1.64) = 0.95$, tenemos
- $P(\hat{\mu} - 1.64 \frac{\sigma}{\sqrt{N}} < \mu) = 0.95$
- Entonces al intervalo $(\hat{\mu} - 1.64 \frac{\sigma}{\sqrt{N}}, \infty)$ se lo llama *intervalo de confianza de 95% superior unilateral*
- Podemos definir análogamente el intervalo de confianza de 95% inferior unilateral como $\mu \in (-\infty, \hat{\mu} + 1.64 \frac{\sigma}{\sqrt{N}})$
- El intervalo de confianza no nos dice nada acerca de la distribución de μ !



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Intervalo de Confianza Unilateral

- $P(\hat{\mu} - z_{\alpha/2} \frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2} \frac{\sigma}{\sqrt{N}}) = 1 - \alpha$
- El intervalo de confianza porcentual $100 \times (1 - \alpha)$ está dentro de $z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$ unidades de la media muestral $\hat{\mu}$
- Dado que $P(Z < 1.64) = 0.95$, tenemos
- $P(\hat{\mu} - 1.64 \frac{\sigma}{\sqrt{N}} < \mu) = 0.95$
- Entonces al intervalo $(\hat{\mu} - 1.64 \frac{\sigma}{\sqrt{N}}, \infty)$ se lo llama *intervalo de confianza de 95% superior unilateral*
- Podemos definir análogamente el intervalo de confianza de 95% inferior unilateral como $\mu \in (-\infty, \hat{\mu} + 1.64 \frac{\sigma}{\sqrt{N}})$
- El intervalo de confianza no nos dice nada acerca de la distribución de μ !
- De hecho μ puede ser una constante desconocida



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Distribución Chi-Cuadrado

- Cuando la varianza es desconocida, vamos a tener que usar una estimación de la varianza



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Distribución Chi-Cuadrado

- Cuando la varianza es desconocida, vamos a tener que usar una estimación de la varianza
- Si Z_i son v.a. i.i.d $\mathcal{N}(0, 1)$, la distribución de la suma de los cuadrados $X = Z_1^2 + Z_2^2 + \dots + Z_n^2$ es:



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Distribución Chi-Cuadrado

- Cuando la varianza es desconocida, vamos a tener que usar una estimación de la varianza
- Si Z_i son v.a. i.i.d $\mathcal{N}(0, 1)$, la distribución de la suma de los cuadrados $X = Z_1^2 + Z_2^2 + \dots + Z_n^2$ es:
- $X \sim \chi_n^2$, donde χ_n^2 es una distribución chi-cuadrado de orden n



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Distribución Chi-Cuadrado

- Cuando la varianza es desconocida, vamos a tener que usar una estimación de la varianza
- Si Z_i son v.a. i.i.d $\mathcal{N}(0, 1)$, la distribución de la suma de los cuadrados $X = Z_1^2 + Z_2^2 + \dots + Z_n^2$ es:
- $X \sim \chi_n^2$, donde χ_n^2 es una distribución chi-cuadrado de orden n
- La pdf es:



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Distribución Chi-Cuadrado

- Cuando la varianza es desconocida, vamos a tener que usar una estimación de la varianza
- Si Z_i son v.a. i.i.d $\mathcal{N}(0, 1)$, la distribución de la suma de los cuadrados $X = Z_1^2 + Z_2^2 + \dots + Z_n^2$ es:
- $X \sim \chi_n^2$, donde χ_n^2 es una distribución chi-cuadrado de orden n
- La pdf es:
- $f_X(x, n) = \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})}$ si $x > 0$, o $f_X(x, n) = 0$ si no



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Distribución Chi-Cuadrado

- Cuando la varianza es desconocida, vamos a tener que usar una estimación de la varianza
- Si Z_i son v.a. i.i.d $\mathcal{N}(0, 1)$, la distribución de la suma de los cuadrados $X = Z_1^2 + Z_2^2 + \dots + Z_n^2$ es:
- $X \sim \chi_n^2$, donde χ_n^2 es una distribución chi-cuadrado de orden n
- La pdf es:
- $f_X(x, n) = \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})}$ si $x > 0$, o $f_X(x, n) = 0$ si no
- $E[X] = n$, $\text{var}[X] = 2n$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Distribución Chi-Cuadrado

- Cuando la varianza es desconocida, vamos a tener que usar una estimación de la varianza
- Si Z_i son v.a. i.i.d $\mathcal{N}(0, 1)$, la distribución de la suma de los cuadrados $X = Z_1^2 + Z_2^2 + \dots + Z_n^2$ es:
- $X \sim \chi_n^2$, donde χ_n^2 es una distribución chi-cuadrado de orden n
- La pdf es:
- $f_X(x, n) = \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})}$ si $x > 0$, o $f_X(x, n) = 0$ si no
- $E[X] = n$, $\text{var}[X] = 2n$
- La suma de cuadrados de $\sum_{i=1}^n (Z_i - \bar{Z})^2$ es una χ_{n-1}^2 , donde $\bar{Z} = (1/n) \sum_{i=1}^n Z_i$ es la media muestral



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Distribución t de Student

- Si $Z \sim \mathcal{N}(0, 1)$ y $X \sim \chi_n^2$ son v.a. independientes entonces el cociente $T_n = \frac{Z}{\sqrt{X/n}}$ tiene distribución t con n grados de libertad, t_n



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Distribución t de Student

- Si $Z \sim \mathcal{N}(0, 1)$ y $X \sim \chi_n^2$ son v.a. independientes entonces el cociente $T_n = \frac{Z}{\sqrt{X/n}}$ tiene distribución t con n grados de libertad, t_n
- La pdf es



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Distribución t de Student

- Si $Z \sim \mathcal{N}(0, 1)$ y $X \sim \chi_n^2$ son v.a. independientes entonces el cociente $T_n = \frac{Z}{\sqrt{X/n}}$ tiene distribución t con n grados de libertad, t_n
- La pdf es
- $$f_T(t, n) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Distribución t de Student

- Si $Z \sim \mathcal{N}(0, 1)$ y $X \sim \chi_n^2$ son v.a. independientes entonces el cociente $T_n = \frac{Z}{\sqrt{X/n}}$ tiene distribución t con n grados de libertad, t_n
- La pdf es
- $f_T(t, n) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$
- $E[T_n] = 0, n > 1$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Distribución t de Student

- Si $Z \sim \mathcal{N}(0, 1)$ y $X \sim \chi_n^2$ son v.a. independientes entonces el cociente $T_n = \frac{Z}{\sqrt{X/n}}$ tiene distribución t con n grados de libertad, t_n
- La pdf es
- $f_T(t, n) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$
- $E[T_n] = 0, n > 1$
- $\text{var}[T_n] = \frac{n}{n-2}$ para $n > 2$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Distribución t de Student

- Si $Z \sim \mathcal{N}(0, 1)$ y $X \sim \chi_n^2$ son v.a. independientes entonces el cociente $T_n = \frac{Z}{\sqrt{X/n}}$ tiene distribución t con n grados de libertad, t_n
- La pdf es
- $f_T(t, n) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$
- $E[T_n] = 0, n > 1$
- $\text{var}[T_n] = \frac{n}{n-2}$ para $n > 2$
- T_n se parece a $\mathcal{N}(0, 1)$ cuando $n \rightarrow \infty$ pero con colas más notorias (mayor variabilidad que la normal cero-uno)



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo con Varianza Desconocida

- Hasta ahora, hemos asumido que conocemos exactamente la varianza σ^2 de las v.a.



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo con Varianza Desconocida

- Hasta ahora, hemos asumido que conocemos exactamente la varianza σ^2 de las v.a.
- En general, la varianza es desconocida



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo con Varianza Desconocida

- Hasta ahora, hemos asumido que conocemos exactamente la varianza σ^2 de las v.a.
- En general, la varianza es desconocida
- Podemos estimar la varianza con el estimador insesgado:



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo con Varianza Desconocida

- Hasta ahora, hemos asumido que conocemos exactamente la varianza σ^2 de las v.a.
- En general, la varianza es desconocida
- Podemos estimar la varianza con el estimador insesgado:
- $S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{\mu})^2$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo con Varianza Desconocida

- Hasta ahora, hemos asumido que conocemos exactamente la varianza σ^2 de las v.a.
- En general, la varianza es desconocida
- Podemos estimar la varianza con el estimador insesgado:
- $S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{\mu})^2$
- La variable $(N-1)S^2/\sigma^2 \sim \chi_{n-1}^2$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo con Varianza Desconocida

- Hasta ahora, hemos asumido que conocemos exactamente la varianza σ^2 de las v.a.
- En general, la varianza es desconocida
- Podemos estimar la varianza con el estimador insesgado:
- $S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{\mu})^2$
- La variable $(N-1)S^2/\sigma^2 \sim \chi_{n-1}^2$
- Además, $\hat{\mu}, S^2$ son independientes



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo con Varianza Desconocida

- Hasta ahora, hemos asumido que conocemos exactamente la varianza σ^2 de las v.a.
- En general, la varianza es desconocida
- Podemos estimar la varianza con el estimador insesgado:
- $S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{\mu})^2$
- La variable $(N-1)S^2/\sigma^2 \sim \chi_{n-1}^2$
- Además, $\hat{\mu}, S^2$ son independientes
- Entonces, la variable $\sqrt{N}(\hat{\mu} - \mu)/S \sim t_{N-1}$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo con Varianza Desconocida

- Entonces, ahora podemos estimar el intervalo de confianza con varianza desconocida usando la distribución t en lugar de la normal cero-uno



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo con Varianza Desconocida

- Entonces, ahora podemos estimar el intervalo de confianza con varianza desconocida usando la distribución t en lugar de la normal cero-uno
- $P(t_{1-\alpha/2, N-1} < \sqrt{N} \frac{(\hat{\mu} - \mu)}{S} < t_{\alpha/2, N-1}) = 1 - \alpha$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo con Varianza Desconocida

- Entonces, ahora podemos estimar el intervalo de confianza con varianza desconocida usando la distribución t en lugar de la normal cero-uno
- $P(t_{1-\alpha/2, N-1} < \sqrt{N} \frac{(\hat{\mu} - \mu)}{S} < t_{\alpha/2, N-1}) = 1 - \alpha$
- Usando $t_{1-\alpha/2, N-1} = -t_{\alpha/2, N-1}$, y despejando μ obtenemos



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo con Varianza Desconocida

- Entonces, ahora podemos estimar el intervalo de confianza con varianza desconocida usando la distribución t en lugar de la normal cero-uno
- $P(t_{1-\alpha/2, N-1} < \sqrt{N} \frac{(\hat{\mu} - \mu)}{S} < t_{\alpha/2, N-1}) = 1 - \alpha$
- Usando $t_{1-\alpha/2, N-1} = -t_{\alpha/2, N-1}$, y despejando μ obtenemos
- $P(\hat{\mu} - t_{\alpha/2, N-1} \frac{S}{\sqrt{N}} < \mu < \hat{\mu} + t_{\alpha/2, N-1} \frac{S}{\sqrt{N}}) = 1 - \alpha$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo con Varianza Desconocida

- Entonces, ahora podemos estimar el intervalo de confianza con varianza desconocida usando la distribución t en lugar de la normal cero-uno
- $P(t_{1-\alpha/2, N-1} < \sqrt{N} \frac{(\hat{\mu} - \mu)}{S} < t_{\alpha/2, N-1}) = 1 - \alpha$
- Usando $t_{1-\alpha/2, N-1} = -t_{\alpha/2, N-1}$, y despejando μ obtenemos
- $P(\hat{\mu} - t_{\alpha/2, N-1} \frac{S}{\sqrt{N}} < \mu < \hat{\mu} + t_{\alpha/2, N-1} \frac{S}{\sqrt{N}}) = 1 - \alpha$
- Nota: para N grande, podemos usar la normal cero-uno Z en lugar de la T



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Estimación de Intervalo con Varianza Desconocida

- Entonces, ahora podemos estimar el intervalo de confianza con varianza desconocida usando la distribución t en lugar de la normal cero-uno
- $P(t_{1-\alpha/2, N-1} < \sqrt{N} \frac{(\hat{\mu} - \mu)}{S} < t_{\alpha/2, N-1}) = 1 - \alpha$
- Usando $t_{1-\alpha/2, N-1} = -t_{\alpha/2, N-1}$, y despejando μ obtenemos
- $P(\hat{\mu} - t_{\alpha/2, N-1} \frac{S}{\sqrt{N}} < \mu < \hat{\mu} + t_{\alpha/2, N-1} \frac{S}{\sqrt{N}}) = 1 - \alpha$
- Nota: para N grande, podemos usar la normal cero-uno Z en lugar de la T
- Cuando el número de muestras es chico, usamos la t de Student y *pagamos* el precio de usar un intervalo más grande para el mismo nivel de confianza



**FACULTAD
DE INGENIERÍA**
Universidad de Buenos Aires

Ejemplo de Intervalo de Confianza

- De una muestra de 200 puntos de una v.a. obtenemos que su media muestral es 2.49 y su desviación estándar muestral es 0.674.



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Ejemplo de Intervalo de Confianza

- De una muestra de 200 puntos de una v.a. obtenemos que su media muestral es 2.49 y su desviación estándar muestral es 0.674.
- Encontrar un intervalo de confianza del 96% de la media



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Ejemplo de Intervalo de Confianza

- De una muestra de 200 puntos de una v.a. obtenemos que su media muestral es 2.49 y su desviación estándar muestral es 0.674.
- Encontrar un intervalo de confianza del 96% de la media
- Solución:



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Ejemplo de Intervalo de Confianza

- De una muestra de 200 puntos de una v.a. obtenemos que su media muestral es 2.49 y su desviación estándar muestral es 0.674.
- Encontrar un intervalo de confianza del 96% de la media
- Solución:
- Estamos lidiando con una varianza estimada, entonces usamos la t de Student



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Ejemplo de Intervalo de Confianza

- De una muestra de 200 puntos de una v.a. obtenemos que su media muestral es 2.49 y su desviación estándar muestral es 0.674.
- Encontrar un intervalo de confianza del 96% de la media
- Solución:
- Estamos lidiando con una varianza estimada, entonces usamos la t de Student
- Primero buscamos el valor $t_{1-\alpha/2, N-1} = -t_{\alpha/2, N-1}$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Ejemplo de Intervalo de Confianza

- De una muestra de 200 puntos de una v.a. obtenemos que su media muestral es 2.49 y su desviación estándar muestral es 0.674.
- Encontrar un intervalo de confianza del 96% de la media
- Solución:
- Estamos lidiando con una varianza estimada, entonces usamos la t de Student
- Primero buscamos el valor $t_{1-\alpha/2, N-1} = -t_{\alpha/2, N-1}$
- Vemos que $\alpha = 1 - 0.96 = 0.04$, y que $N = 200$, entonces nos interesa encontrar $t_{1-(0.04/2), (200-1)} = t_{0.98, 199}$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Ejemplo de Intervalo de Confianza

- De una muestra de 200 puntos de una v.a. obtenemos que su media muestral es 2.49 y su desviación estándar muestral es 0.674.
- Encontrar un intervalo de confianza del 96% de la media
- Solución:
- Estamos lidiando con una varianza estimada, entonces usamos la t de Student
- Primero buscamos el valor $t_{1-\alpha/2, N-1} = -t_{\alpha/2, N-1}$
- Vemos que $\alpha = 1 - 0.96 = 0.04$, y que $N = 200$, entonces nos interesa encontrar $t_{1-(0.04/2), (200-1)} = t_{0.98, 199}$
- En Octave, usamos la función `tinu` para encontrar el cuantil



Ejemplo de Intervalo de Confianza

- De una muestra de 200 puntos de una v.a. obtenemos que su media muestral es 2.49 y su desviación estándar muestral es 0.674.
- Encontrar un intervalo de confianza del 96% de la media
- Solución:
- Estamos lidiando con una varianza estimada, entonces usamos la t de Student
- Primero buscamos el valor $t_{1-\alpha/2, N-1} = -t_{\alpha/2, N-1}$
- Vemos que $\alpha = 1 - 0.96 = 0.04$, y que $N = 200$, entonces nos interesa encontrar $t_{1-(0.04/2), (200-1)} = t_{0.98, 199}$
- En Octave, usamos la función `tinu` para encontrar el cuantil
- Esto nos da $t_{0.98, 199} = 2.0673$



Ejemplo de Intervalo de Confianza

- De una muestra de 200 puntos de una v.a. obtenemos que su media muestral es 2.49 y su desviación estándar muestral es 0.674.
- Encontrar un intervalo de confianza del 96% de la media
- Solución:
- Estamos lidiando con una varianza estimada, entonces usamos la t de Student
- Primero buscamos el valor $t_{1-\alpha/2, N-1} = -t_{\alpha/2, N-1}$
- Vemos que $\alpha = 1 - 0.96 = 0.04$, y que $N = 200$, entonces nos interesa encontrar $t_{1-(0.04/2), (200-1)} = t_{0.98, 199}$
- En Octave, usamos la función `tinu` para encontrar el cuantil
- Esto nos da $t_{0.98, 199} = 2.0673$
- El intervalo de confianza del 96% de μ es
$$\mu \in (2.49 \pm 2.0673 \times \frac{0.674}{\sqrt{200}}) = (2.49 \pm 0.099)$$



Test de Hipótesis

- A veces vamos a querer ensayar una hipótesis



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Test de Hipótesis

- A veces vamos a querer ensayar una hipótesis
- Por ejemplo si la media es mayor o distinta de un valor



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Test de Hipótesis

- A veces vamos a querer ensayar una hipótesis
- Por ejemplo si la media es mayor o distinta de un valor
- Una hipótesis es el planteo de una pregunta científica en términos de un valor propuesto de un parámetro en un modelo probabilístico



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Test de Hipótesis

- A veces vamos a querer ensayar una hipótesis
- Por ejemplo si la media es mayor o distinta de un valor
- Una hipótesis es el planteo de una pregunta científica en términos de un valor propuesto de un parámetro en un modelo probabilístico
- Si las muestras aleatorias son consistentes con la hipótesis, aceptamos la hipótesis



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Test de Hipótesis

- A veces vamos a querer ensayar una hipótesis
- Por ejemplo si la media es mayor o distinta de un valor
- Una hipótesis es el planteo de una pregunta científica en términos de un valor propuesto de un parámetro en un modelo probabilístico
- Si las muestras aleatorias son consistentes con la hipótesis, aceptamos la hipótesis
- N.B.: varios textos hablan de *no rechazar* en lugar de *aceptar* la hipótesis



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Test de Hipótesis

- A veces vamos a querer ensayar una hipótesis
- Por ejemplo si la media es mayor o distinta de un valor
- Una hipótesis es el planteo de una pregunta científica en términos de un valor propuesto de un parámetro en un modelo probabilístico
- Si las muestras aleatorias son consistentes con la hipótesis, aceptamos la hipótesis
- N.B.: varios textos hablan de *no rechazar* en lugar de *aceptar* la hipótesis
- En caso contrario, rechazamos la hipótesis



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Test de Hipótesis

- A veces vamos a querer ensayar una hipótesis
- Por ejemplo si la media es mayor o distinta de un valor
- Una hipótesis es el planteo de una pregunta científica en términos de un valor propuesto de un parámetro en un modelo probabilístico
- Si las muestras aleatorias son consistentes con la hipótesis, aceptamos la hipótesis
- N.B.: varios textos hablan de *no rechazar* en lugar de *aceptar* la hipótesis
- En caso contrario, rechazamos la hipótesis
- De todas formas, nunca decimos que la hipótesis es verdadera o falsa, sino que los datos la avalan o no con un cierto nivel de confianza



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Test de Hipótesis

- En un test de hipótesis definimos una estadística que debe obedecer cierta distribución si la hipótesis es correcta



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Test de Hipótesis

- En un test de hipótesis definimos una estadística que debe obedecer cierta distribución si la hipótesis es correcta
- Si la estadística calculada tiene muy baja probabilidad de ocurrir bajo esa hipótesis, rechazamos la hipótesis



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Test de Hipótesis

- En un test de hipótesis definimos una estadística que debe obedecer cierta distribución si la hipótesis es correcta
- Si la estadística calculada tiene muy baja probabilidad de ocurrir bajo esa hipótesis, rechazamos la hipótesis
- Caso contrario, la aceptamos



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Test de Hipótesis

- En un test de hipótesis definimos una estadística que debe obedecer cierta distribución si la hipótesis es correcta
- Si la estadística calculada tiene muy baja probabilidad de ocurrir bajo esa hipótesis, rechazamos la hipótesis
- Caso contrario, la aceptamos
- Por ejemplo, tenemos muestras de una distribución de la que desconocemos la media μ pero conocemos la varianza σ^2



**FACULTAD
DE INGENIERÍA**
Universidad de Buenos Aires

Test de Hipótesis

- En un test de hipótesis definimos una estadística que debe obedecer cierta distribución si la hipótesis es correcta
- Si la estadística calculada tiene muy baja probabilidad de ocurrir bajo esa hipótesis, rechazamos la hipótesis
- Caso contrario, la aceptamos
- Por ejemplo, tenemos muestras de una distribución de la que desconocemos la media μ pero conocemos la varianza σ^2
- Queremos ensayar la hipótesis que $\mu = \mu_0$



**FACULTAD
DE INGENIERÍA**
Universidad de Buenos Aires

Test de Hipótesis

- En un test de hipótesis definimos una estadística que debe obedecer cierta distribución si la hipótesis es correcta
- Si la estadística calculada tiene muy baja probabilidad de ocurrir bajo esa hipótesis, rechazamos la hipótesis
- Caso contrario, la aceptamos
- Por ejemplo, tenemos muestras de una distribución de la que desconocemos la media μ pero conocemos la varianza σ^2
- Queremos ensayar la hipótesis que $\mu = \mu_0$
- La hipótesis nula es $\mathcal{H}_0 : \mu = \mu_0$



**FACULTAD
DE INGENIERIA**

Universidad de Buenos Aires

Test de Hipótesis

- En un test de hipótesis definimos una estadística que debe obedecer cierta distribución si la hipótesis es correcta
- Si la estadística calculada tiene muy baja probabilidad de ocurrir bajo esa hipótesis, rechazamos la hipótesis
- Caso contrario, la aceptamos
- Por ejemplo, tenemos muestras de una distribución de la que desconocemos la media μ pero conocemos la varianza σ^2
- Queremos ensayar la hipótesis que $\mu = \mu_0$
- La hipótesis nula es $\mathcal{H}_0 : \mu = \mu_0$
- La hipótesis alternativa es $\mathcal{H}_1 : \mu = \mu_1 \neq \mu_0$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

- Vamos a comparar la media muestral con la hipótesis



Test de Hipótesis

- Vamos a comparar la media muestral con la hipótesis
- Si la media muestral cae fuera de un intervalo de confianza $100 \times (1 - \alpha)$ rechazamos la hipótesis nula



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Test de Hipótesis

- Vamos a comparar la media muestral con la hipótesis
- Si la media muestral cae fuera de un intervalo de confianza $100 \times (1 - \alpha)$ rechazamos la hipótesis nula
- Es decir:



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Test de Hipótesis

- Vamos a comparar la media muestral con la hipótesis
- Si la media muestral cae fuera de un intervalo de confianza $100 \times (1 - \alpha)$ rechazamos la hipótesis nula
- Es decir:
- Aceptamos la hipótesis si $\frac{\sqrt{N}(\hat{\mu} - \mu)}{\sigma} \in (-z_{\alpha/2}, z_{\alpha/2})$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Test de Hipótesis

- Vamos a comparar la media muestral con la hipótesis
- Si la media muestral cae fuera de un intervalo de confianza $100 \times (1 - \alpha)$ rechazamos la hipótesis nula
- Es decir:
- Aceptamos la hipótesis si $\frac{\sqrt{N}(\hat{\mu} - \mu)}{\sigma} \in (-z_{\alpha/2}, z_{\alpha/2})$
- Este es un test bilateral (rechazamos la hipótesis si μ cae fuera del intervalo)



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

- Error de tipo I: ocurre cuando rechazamos la hipótesis cuando es correcta, es decir es una *falsa alarma*



Tipos de Errores

- Error de tipo I: ocurre cuando rechazamos la hipótesis cuando es correcta, es decir es una *falsa alarma*
- Formalmente, $\alpha = P(\text{Tipo I} | \mathcal{H}_0)$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Tipos de Errores

- Error de tipo I: ocurre cuando rechazamos la hipótesis cuando es correcta, es decir es una *falsa alarma*
- Formalmente, $\alpha = P(\text{Tipo I} | \mathcal{H}_0)$
- La p. de error tipo I es α , típicamente se elige 0.1, 0.05, 0.01



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Tipos de Errores

- Error de tipo I: ocurre cuando rechazamos la hipótesis cuando es correcta, es decir es una *falsa alarma*
- Formalmente, $\alpha = P(\text{Tipo I} | \mathcal{H}_0)$
- La p. de error tipo I es α , típicamente se elige 0.1, 0.05, 0.01
- Error de tipo II: ocurre cuando aceptamos la hipótesis cuando es incorrecta, es decir *no activar la alarma*



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Tipos de Errores

- Error de tipo I: ocurre cuando rechazamos la hipótesis cuando es correcta, es decir es una *falsa alarma*
- Formalmente, $\alpha = P(\text{Tipo I}|\mathcal{H}_0)$
- La p. de error tipo I es α , típicamente se elige 0.1, 0.05, 0.01
- Error de tipo II: ocurre cuando aceptamos la hipótesis cuando es incorrecta, es decir *no activar la alarma*
- $\beta = P(\text{Tipo II}|\mathcal{H}_1)$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Tipos de Errores

- Error de tipo I: ocurre cuando rechazamos la hipótesis cuando es correcta, es decir es una *falsa alarma*
- Formalmente, $\alpha = P(\text{Tipo I}|\mathcal{H}_0)$
- La p. de error tipo I es α , típicamente se elige 0.1, 0.05, 0.01
- Error de tipo II: ocurre cuando aceptamos la hipótesis cuando es incorrecta, es decir *no activar la alarma*
- $\beta = P(\text{Tipo II}|\mathcal{H}_1)$
- $\beta(\mu) = P_{\mu} \left(-z_{\alpha/2} < \frac{\sqrt{N}(\hat{\mu} - \mu_0)}{\sigma} < z_{\alpha/2} \right)$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Tipos de Errores

- Error de tipo I: ocurre cuando rechazamos la hipótesis cuando es correcta, es decir es una *falsa alarma*
- Formalmente, $\alpha = P(\text{Tipo I}|\mathcal{H}_0)$
- La p. de error tipo I es α , típicamente se elige 0.1, 0.05, 0.01
- Error de tipo II: ocurre cuando aceptamos la hipótesis cuando es incorrecta, es decir *no activar la alarma*
- $\beta = P(\text{Tipo II}|\mathcal{H}_1)$
- $\beta(\mu) = P_{\mu} \left(-z_{\alpha/2} < \frac{\sqrt{N}(\hat{\mu} - \mu_0)}{\sigma} < z_{\alpha/2} \right)$
- Esta p. aumenta a medida que μ se acerca a μ_0



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Poder del Ensayo

- Se define a $Poder(\mu) \triangleq 1 - \beta(\mu)$ como la función de poder del ensayo, es decir la p. de rechazar cuando el valor real es distinto de μ_0 (ensayo bilateral)



**FACULTAD
DE INGENIERIA**

Universidad de Buenos Aires

Poder del Ensayo

- Se define a $Poder(\mu) \triangleq 1 - \beta(\mu)$ como la función de poder del ensayo, es decir la p. de rechazar cuando el valor real es distinto de μ_0 (ensayo bilateral)
- Podemos elegir el tamaño de la muestra N de manera de controlar el poder del ensayo $1 - \beta(\mu)$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Poder del Ensayo

- Se define a $Poder(\mu) \triangleq 1 - \beta(\mu)$ como la función de poder del ensayo, es decir la p. de rechazar cuando el valor real es distinto de μ_0 (ensayo bilateral)
- Podemos elegir el tamaño de la muestra N de manera de controlar el poder del ensayo $1 - \beta(\mu)$
- $Poder(\mu) = P_{\mu}(\hat{\mu} > \mu_0 + z_{1-\alpha/2} \sigma / \sqrt{N}) + P_{\mu}(\hat{\mu} < \mu_0 - z_{1-\alpha/2} \sigma / \sqrt{N})$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Poder del Ensayo

- Se define a $Poder(\mu) \triangleq 1 - \beta(\mu)$ como la función de poder del ensayo, es decir la p. de rechazar cuando el valor real es distinto de μ_0 (ensayo bilateral)
- Podemos elegir el tamaño de la muestra N de manera de controlar el poder del ensayo $1 - \beta(\mu)$
- $Poder(\mu) = P_{\mu}(\hat{\mu} > \mu_0 + z_{1-\alpha/2} \sigma / \sqrt{N}) + P_{\mu}(\hat{\mu} < \mu_0 - z_{1-\alpha/2} \sigma / \sqrt{N})$
- $Poder(\mu) = P(\frac{\hat{\mu} - \mu}{\sigma / \sqrt{N}} > \frac{\mu_0 - \mu}{\sigma / \sqrt{N}} + z_{1-\alpha/2}) + P(\frac{\hat{\mu} - \mu}{\sigma / \sqrt{N}} < \frac{\mu_0 - \mu}{\sigma / \sqrt{N}} - z_{1-\alpha/2})$



Poder del Ensayo

- Se define a $Poder(\mu) \triangleq 1 - \beta(\mu)$ como la función de poder del ensayo, es decir la p. de rechazar cuando el valor real es distinto de μ_0 (ensayo bilateral)
- Podemos elegir el tamaño de la muestra N de manera de controlar el poder del ensayo $1 - \beta(\mu)$
- $Poder(\mu) = P_{\mu}(\hat{\mu} > \mu_0 + z_{1-\alpha/2} \sigma / \sqrt{N}) + P_{\mu}(\hat{\mu} < \mu_0 - z_{1-\alpha/2} \sigma / \sqrt{N})$
- $Poder(\mu) = P(\frac{\hat{\mu} - \mu}{\sigma / \sqrt{N}} > \frac{\mu_0 - \mu}{\sigma / \sqrt{N}} + z_{1-\alpha/2}) + P(\frac{\hat{\mu} - \mu}{\sigma / \sqrt{N}} < \frac{\mu_0 - \mu}{\sigma / \sqrt{N}} - z_{1-\alpha/2})$
- Definimos $\Delta \triangleq \frac{\mu - \mu_0}{\sigma / \sqrt{N}}$, y $\Phi(z) = P(Z \leq z)$, con $Z \sim \mathcal{N}(0, 1)$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Poder del Ensayo

- Se define a $Poder(\mu) \triangleq 1 - \beta(\mu)$ como la función de poder del ensayo, es decir la p. de rechazar cuando el valor real es distinto de μ_0 (ensayo bilateral)
- Podemos elegir el tamaño de la muestra N de manera de controlar el poder del ensayo $1 - \beta(\mu)$
- $Poder(\mu) = P_{\mu}(\hat{\mu} > \mu_0 + z_{1-\alpha/2} \sigma / \sqrt{N}) + P_{\mu}(\hat{\mu} < \mu_0 - z_{1-\alpha/2} \sigma / \sqrt{N})$
- $Poder(\mu) = P(\frac{\hat{\mu} - \mu}{\sigma / \sqrt{N}} > \frac{\mu_0 - \mu}{\sigma / \sqrt{N}} + z_{1-\alpha/2}) + P(\frac{\hat{\mu} - \mu}{\sigma / \sqrt{N}} < \frac{\mu_0 - \mu}{\sigma / \sqrt{N}} - z_{1-\alpha/2})$
- Definimos $\Delta \triangleq \frac{\mu - \mu_0}{\sigma / \sqrt{N}}$, y $\Phi(z) = P(Z \leq z)$, con $Z \sim \mathcal{N}(0, 1)$
- $Poder(\mu) = \Phi(\sqrt{N}\Delta - z_{1-\alpha/2}) + \Phi(-\sqrt{N}\Delta - z_{1-\alpha/2})$



Poder del Ensayo

- Se define a $Poder(\mu) \triangleq 1 - \beta(\mu)$ como la función de poder del ensayo, es decir la p. de rechazar cuando el valor real es distinto de μ_0 (ensayo bilateral)
- Podemos elegir el tamaño de la muestra N de manera de controlar el poder del ensayo $1 - \beta(\mu)$
- $Poder(\mu) = P_{\mu}(\hat{\mu} > \mu_0 + z_{1-\alpha/2} \sigma / \sqrt{N}) + P_{\mu}(\hat{\mu} < \mu_0 - z_{1-\alpha/2} \sigma / \sqrt{N})$
- $Poder(\mu) = P(\frac{\hat{\mu} - \mu}{\sigma / \sqrt{N}} > \frac{\mu_0 - \mu}{\sigma / \sqrt{N}} + z_{1-\alpha/2}) + P(\frac{\hat{\mu} - \mu}{\sigma / \sqrt{N}} < \frac{\mu_0 - \mu}{\sigma / \sqrt{N}} - z_{1-\alpha/2})$
- Definimos $\Delta \triangleq \frac{\mu - \mu_0}{\sigma / \sqrt{N}}$, y $\Phi(z) = P(Z \leq z)$, con $Z \sim \mathcal{N}(0, 1)$
- $Poder(\mu) = \Phi(\sqrt{N}\Delta - z_{1-\alpha/2}) + \Phi(-\sqrt{N}\Delta - z_{1-\alpha/2})$
- Se aproxima $\Phi(-\sqrt{N}\Delta - z_{1-\alpha/2}) \approx 0$ si $\sqrt{N}\Delta > 1$, entonces:



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Poder del Ensayo

- Se define a $Poder(\mu) \triangleq 1 - \beta(\mu)$ como la función de poder del ensayo, es decir la p. de rechazar cuando el valor real es distinto de μ_0 (ensayo bilateral)
- Podemos elegir el tamaño de la muestra N de manera de controlar el poder del ensayo $1 - \beta(\mu)$
- $Poder(\mu) = P_{\mu}(\hat{\mu} > \mu_0 + z_{1-\alpha/2} \sigma / \sqrt{N}) + P_{\mu}(\hat{\mu} < \mu_0 - z_{1-\alpha/2} \sigma / \sqrt{N})$
- $Poder(\mu) = P(\frac{\hat{\mu} - \mu}{\sigma / \sqrt{N}} > \frac{\mu_0 - \mu}{\sigma / \sqrt{N}} + z_{1-\alpha/2}) + P(\frac{\hat{\mu} - \mu}{\sigma / \sqrt{N}} < \frac{\mu_0 - \mu}{\sigma / \sqrt{N}} - z_{1-\alpha/2})$
- Definimos $\Delta \triangleq \frac{\mu - \mu_0}{\sigma / \sqrt{N}}$, y $\Phi(z) = P(Z \leq z)$, con $Z \sim \mathcal{N}(0, 1)$
- $Poder(\mu) = \Phi(\sqrt{N}\Delta - z_{1-\alpha/2}) + \Phi(-\sqrt{N}\Delta - z_{1-\alpha/2})$
- Se aproxima $\Phi(-\sqrt{N}\Delta - z_{1-\alpha/2}) \approx 0$ si $\sqrt{N}\Delta > 1$, entonces:
- $Poder(\mu) = \Phi(\sqrt{N}\Delta - z_{1-\alpha/2})$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

- Si queremos configurar o prescribir $Poder(\mu) \triangleq 1 - \beta(\mu) = 0.9$ (por ej.), podemos elegir N



Poder del Ensayo y Tamaño de la Muestra

- Si queremos configurar o prescribir $Poder(\mu) \triangleq 1 - \beta(\mu) = 0.9$ (por ej.), podemos elegir N
- Tenemos que resolver la siguiente ecuación:



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Poder del Ensayo y Tamaño de la Muestra

- Si queremos configurar o prescribir $Poder(\mu) \triangleq 1 - \beta(\mu) = 0.9$ (por ej.), podemos elegir N
- Tenemos que resolver la siguiente ecuación:
- $\Phi(\sqrt{N}\Delta - z_{1-\alpha/2}) = 1 - \beta$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Poder del Ensayo y Tamaño de la Muestra

- Si queremos configurar o prescribir $\text{Poder}(\mu) \triangleq 1 - \beta(\mu) = 0.9$ (por ej.), podemos elegir N
- Tenemos que resolver la siguiente ecuación:
- $\Phi(\sqrt{N}\Delta - z_{1-\alpha/2}) = 1 - \beta$
- Entonces, N es



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Poder del Ensayo y Tamaño de la Muestra

- Si queremos configurar o prescribir $Poder(\mu) \triangleq 1 - \beta(\mu) = 0.9$ (por ej.), podemos elegir N
- Tenemos que resolver la siguiente ecuación:
- $\Phi(\sqrt{N}\Delta - z_{1-\alpha/2}) = 1 - \beta$
- Entonces, N es
- $N = (z_{1-\alpha/2} + z_{1-\beta})^2 \frac{1}{\Delta^2}$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

- Para un ensayo unilateral, tenemos:



Ensayo Unilateral

- Para un ensayo unilateral, tenemos:
- $\mathcal{H}_0 : \mu \leq \mu_0$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Ensayo Unilateral

- Para un ensayo unilateral, tenemos:
- $\mathcal{H}_0 : \mu \leq \mu_0$
- $\mathcal{H}_1 : \mu > \mu_0$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Ensayo Unilateral

- Para un ensayo unilateral, tenemos:
- $\mathcal{H}_0 : \mu \leq \mu_0$
- $\mathcal{H}_1 : \mu > \mu_0$
- El intervalo de confianza de nivel $100 \times (1 - \alpha)$ ahora se define en base a un solo lado donde $\hat{\mu}$ debe caer para no rechazar \mathcal{H}_0



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

- Para un ensayo unilateral, tenemos:
- $\mathcal{H}_0 : \mu \leq \mu_0$
- $\mathcal{H}_1 : \mu > \mu_0$
- El intervalo de confianza de nivel $100 \times (1 - \alpha)$ ahora se define en base a un solo lado donde $\hat{\mu}$ debe caer para no rechazar \mathcal{H}_0
- Aceptamos \mathcal{H}_0 si:



Ensayo Unilateral

- Para un ensayo unilateral, tenemos:
- $\mathcal{H}_0 : \mu \leq \mu_0$
- $\mathcal{H}_1 : \mu > \mu_0$
- El intervalo de confianza de nivel $100 \times (1 - \alpha)$ ahora se define en base a un solo lado donde $\hat{\mu}$ debe caer para no rechazar \mathcal{H}_0
- Aceptamos \mathcal{H}_0 si:
- $\frac{\sqrt{N}(\hat{\mu} - \mu_0)}{\sigma} \in (-\infty, z_\alpha)$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Test de Hipótesis con Varianza Desconocida

- Si la varianza es desconocida, usamos la distribución t de Student



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Test de Hipótesis con Varianza Desconocida

- Si la varianza es desconocida, usamos la distribución t de Student

- $\frac{\sqrt{N}(\hat{\mu}-\mu)}{\sigma} \sim t_{N-1}$



**FACULTAD
DE INGENIERÍA**
Universidad de Buenos Aires

Test de Hipótesis con Varianza Desconocida

- Si la varianza es desconocida, usamos la distribución t de Student
- $\frac{\sqrt{N}(\hat{\mu}-\mu)}{\sigma} \sim t_{N-1}$
- Para el caso anterior de ensayo bilateral de medias, aceptamos \mathcal{H}_0 si



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Test de Hipótesis con Varianza Desconocida

- Si la varianza es desconocida, usamos la distribución t de Student
- $\frac{\sqrt{N}(\hat{\mu}-\mu)}{\sigma} \sim t_{N-1}$
- Para el caso anterior de ensayo bilateral de medias, aceptamos \mathcal{H}_0 si
- $\frac{\sqrt{N}(\hat{\mu}-\mu)}{\sigma} \in (-t_{\alpha/2,(N-1)}, t_{\alpha/2,(N-1)})$



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Test de Hipótesis con Varianza Desconocida

- Si la varianza es desconocida, usamos la distribución t de Student
- $\frac{\sqrt{N}(\hat{\mu}-\mu)}{\sigma} \sim t_{N-1}$
- Para el caso anterior de ensayo bilateral de medias, aceptamos \mathcal{H}_0 si
- $\frac{\sqrt{N}(\hat{\mu}-\mu)}{\sigma} \in (-t_{\alpha/2,(N-1)}, t_{\alpha/2,(N-1)})$
- Es decir, este es un ensayo bilateral t



Ejercicio 1

- a) Se toma una muestra aleatoria de la altura de 100 personas y se obtiene que la altura promedio es 1.8m. La varianza de la población es $\sigma^2 = 49\text{cm}^2$.
- b) Encontrar un intervalo de confianza de 95% para la altura media de la población, μ .
- c) ¿Cómo cambiaría el desarrollo si se usara la varianza muestral?
- d) Simular el ejercicio en Octave



Ejercicio 2

- a) Un estudio hecho en 10 lamparitas muestra que el tiempo medio de duración en horas es 998.9 hs. Las lamparitas tienen un desvío estándar típico de $\sigma = 80$ hs.
- b) Encontrar el intervalo de confianza de 95% del tiempo medio de vida de estas lamparitas
- c) Encontrar el porcentaje del intervalo de confianza que tiene un rango total de horas de 80 hs. (Ayuda: desplazarse un $\pm\Delta z$ a partir del valor medio $\hat{\mu}$ de manera que $80 = 2\Delta z$).



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

Ejercicio 3

- Ⓐ En un concurso con 10 participantes, cada participante prueba una muestra de 3 vasos de bebida. Dos de los 3 vasos contienen la misma bebida marca X, mientras que el vaso restante contiene la bebida marca Y.
- Ⓑ Queremos determinar si la gente realmente puede discriminar la bebida Y con un nivel de significancia de 5%.
- Ⓒ ¿Cuál es la mínima cantidad de personas que deben identificar la bebida Y para concluir que, en general, la bebida Y es *claramente identificable* con respecto a la bebida X?




Ejercicio 4

- a) Un aviso publicitario asegura que 8 de cada 10 personas prefieren la pasta dental marca Z.
- b) ¿Hay suficiente evidencia para fundamentar esta hipótesis con un 5% de significación? (N.B.: En términos estadísticos, la pregunta correcta sería si el aviso es *significativo*.)



**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires

 E. Alpaydin, *Introduction to machine learning*.
MIT press, 2020.

 “MIT, Brain and Cognitive Sciences, Statistics.” https://ocw.mit.edu/courses/brain-and-cognitive-sciences/9-07-statistics-for-brain-and-cognitive-science-fall-2016/lecture-notes/MIT9_07F16_lec12.pdf.
Accessed: 2020-05-15.

