Verificación de hipótesis paramétricas

Mª Isabel Aguilar, Eugenia Cruces y Bárbara Díaz

UNIVERSIDAD DE MÁLAGA
Departamento de Economía Aplicada (Estadística y Econometría)

Parcialmente financiado a través del PIE13-024 (UMA)

Verificación de hipótesis paramétricas

UNIVERSIDAD DE MÁLAGA

- Introducción
- Conceptos básicos
- Región Crítica Óptima
- Verificación de la media en poblaciones normales
- Verificación de la varianza en poblaciones normales
- Verificación de la diferencia de medias en poblaciones normales
- Verificación de la proporción
- Verificación de la diferencia de proporciones
- Nivel de significación empírico o p-valor

UNIVERSIDAD DE MÁLAGA

- **Introducción**
- Conceptos básicos
- Región Crítica Óptima
- Verificación de la media en poblaciones normales
- Verificación de la varianza en poblaciones normales
- Verificación de la diferencia de medias en poblaciones normales
- Verificación de la proporción
- Verificación de la diferencia de proporciones
- Nivel de significación empírico o p-valor

Introducción

- Primer gran bloque de la inferencia estadística:

  ➡ TEORÍA DE LA ESTIMACIÓN

- Segundo gran bloque de la inferencia estadística:

  ➡ CONTRASTE DE HIPÓTESIS

¿Qué es una Hipótesis?

- Suposición o conjetura planteada en forma de afirmación sobre algún fenómeno, elemento o proceso que tiene lugar en cualquier ámbito
- Parte sustancial del análisis científico. A partir de ella la teoría comienza a tomar forma, incluso cuando la hipótesis es refutada

Introducción

UNIVERSIDAD DE MÁLAGA

¿Qué es una Hipótesis estadística? Suposición o conjetura sobre alguna característica desconocida de una o varias variables aleatorias

Hipótesis paramétrica: sobre algún parámetro θ de la distribución de una variable. Se conoce la forma de f(x)

Hipótesis no paramétrica: sobre cualquier otra característica distinta al valor del parámetro (distribución de la población, aleatoriedad, independencia de dos variables, etc.)

- ✓ Ejemplo1: La media de una variable aleatoria normal es 3

$$\boxed{\text{H: } \mu = 3}$$

- ✓ Ejemplo 2: La distribución de probabilidad de una variable es Poisson de parámetro 3

$$\boxed{\text{H: } X \sim \text{P }(3)}$$

UNIVERSIDAD DE MÁLAGA

Verificación o contraste de hipótesis

- Elección entre dos hipótesis en conflicto → **DECISIÓN**
- Este proceso de decisión usa la teoría de la probabilidad (riesgo), ligada a experimentos con resultados dicotómicos
- Las dos alternativas posibles del experimento son las **HIPÓTESIS** a plantear:
  - **HIPÓTESIS NULA (H_0):** Hipótesis de partida, que se mantendrá como válida a menos que los datos muestrales proporcionen evidencia suficiente en contra
  - **HIPÓTESIS ALTERNATIVA (H_1):** la que está en competencia con la hipótesis nula (engloba las posibilidades alternativas a la misma)

Ejemplo:

- Problema: El dueño de un restaurante en venta afirma que el ingreso medio diario del mismo es de 675 €. Un empresario interesado en adquirirlo duda de esta afirmación

- Traducción a términos estadísticos (Hipótesis a plantear):

$$\begin{cases} H_0 : \mu = 675 & \boxed{=} \\ H_1 : \mu < 675 & \boxed{\neq \;,\; < \;,\; >} \end{cases}$$

El empresario sólo decide invertir si los ingresos no son inferiores a 675 €

Introducción

- **Información** para la toma de decisión: Muestra representativa (*n*)

- **Criterio de decisión** en función de la **discrepancia** entre lo observado y lo propuesto como hipótesis nula (p. ej. $\bar{X} < 625$ → Rechazar H_0)

- **SOLUCIÓN**: Elegir entre H_0 y H_1 en un ambiente de incertidumbre:

Probabilidad → Riesgo → Coste

Introducción

¡¡Son muchos!!

¿Será la media 675?

MUESTRA ALEATORIA

$\bar{x}_{obs}$=600

VAYA ENGAÑO!!!!!

¡¡Gran diferencia!!

Rechazo H_0

- Introducción
- **Conceptos básicos**
- Región Crítica Óptima
- Verificación de la media en poblaciones normales
- Verificación de la varianza en poblaciones normales
- Verificación de la diferencia de medias en poblaciones normales
- Verificación de la proporción
- Verificación de la diferencia de proporciones
- Nivel de significación empírico o p-valor

Conceptos básicos

- **Tipos de hipótesis**

➢ **Simple**: asigna un único valor al parámetro ($\theta = \theta_0$)

$\Longrightarrow$ f(x) queda totalmente especificada

➢ **Compuesta:** establece un intervalo de valores para el parámetro

$\Longrightarrow$ f(x) no queda totalmente especificada

$\theta \neq \theta_0$ (bilateral)

$\theta > \theta_0$ (unilateral derecha)

$\theta < \theta_0$ (unilateral izquierda)

Conceptos básicos

- **Test, contraste o prueba:** regla o criterio de decisión que nos permite decir cuál de las dos hipótesis es más acertada

  Se basa en dos cuestiones:

  - La definición de una **Región Crítica** a partir de un estadístico muestral
  - El valor que toma ese estadístico en la **muestra** seleccionada

Conceptos básicos

- **Región crítica:** Subconjunto del espacio muestral (denotado por C) tal que si valor observado en la muestra pertenece a él, entonces se rechaza la hipótesis nula

$$\boxed{(x_1, x_2, ..., x_n) \in C \rightarrow H_0 \text{ se rechaza}}$$

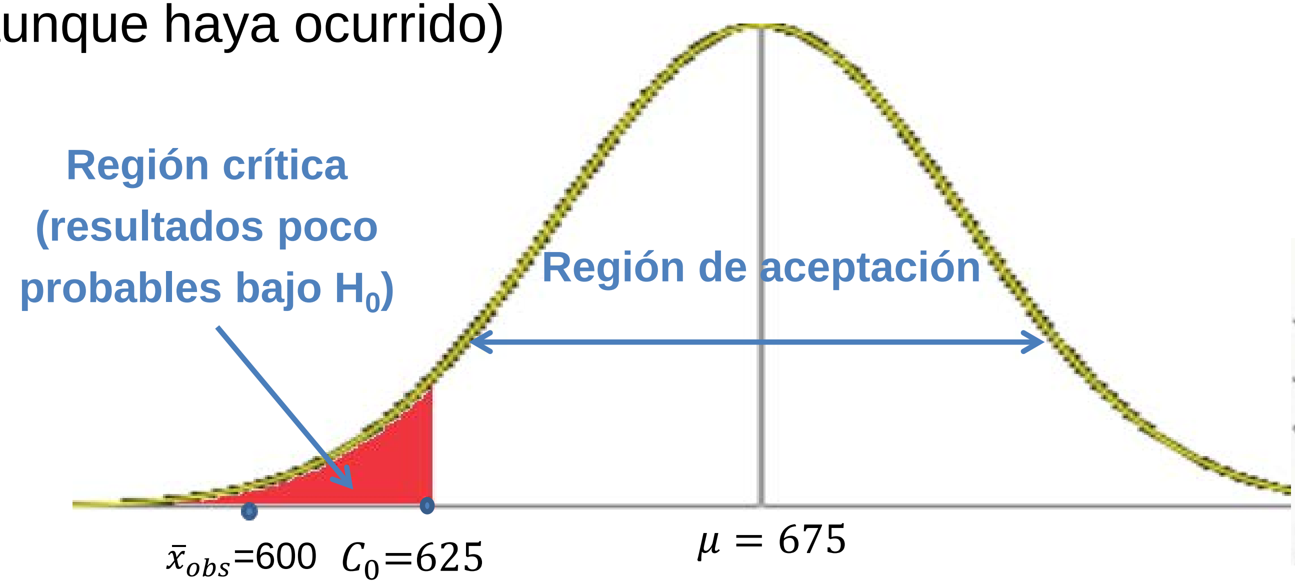
- **Región de aceptación:** Subconjunto complementario de la región crítica (denotado por $\bar{C}$) que nos lleva a aceptar la hipótesis nula

Ejemplo: $C\!: \; \bar{X} > C_0 \longrightarrow \bar{C}\!: \; \bar{X} \leq C_0$

C_0: **valor crítico** que separa la región crítica de la de aceptación

Siguiendo con nuestro ejemplo del restaurante en venta…

Si la afirmación del vendedor fuera cierta ($\mu = 675$), el resultado obtenido en la muestra seleccionada ($\bar{x}_{obs}$=600) sería poco probable (aunque haya ocurrido)



El valor observado de la media muestral pertenece a la región crítica → **Se rechaza H_0** → **El empresario no admite la afirmación del vendedor y decide no comprar**

Conceptos básicos

- **Tipos de error:**

Error de tipo I: se comete al rechazar H_0 *cuando* H_0 es cierta

Error de tipo II: se comete al aceptar H_0 *cuando* H_0 es falsa

Situación real	**Decisión**	
	Rechazar H_0	**Aceptar H_0**
H_0 cierta	Error Tipo I	Decisión correcta
H_0 falsa	Decisión correcta	Error Tipo II

UNIVERSIDAD DE MÁLAGA

- **Tamaño de los errores:** riesgo asumido medido en términos de probabilidad

Situación real	Decisión	
	Rechazar H_0	Aceptar H_0
H_0 cierta	α	$1-\alpha$
H_0 falsa	$1-\beta$	β

Conceptos básicos

UNIVERSIDAD DE MÁLAGA

- **α:** tamaño del error de tipo I; nivel de significación; tamaño de la región crítica

$$\alpha = P(\text{error tipo I}) = P(\text{rechazar } H_0 / H_0 \text{ cierta}) = P[(X_1, X_2 ..., X_n) \in C / H_0]$$

- **β:** tamaño del error de tipo II

$$\beta = P(\text{error tipo II}) = P(\text{aceptar } H_0 / H_0 \text{ falsa}) = P[(X_1, X_2 ..., X_n) \in \overline{C} / H_1]$$

- Probabilidades complementarias:

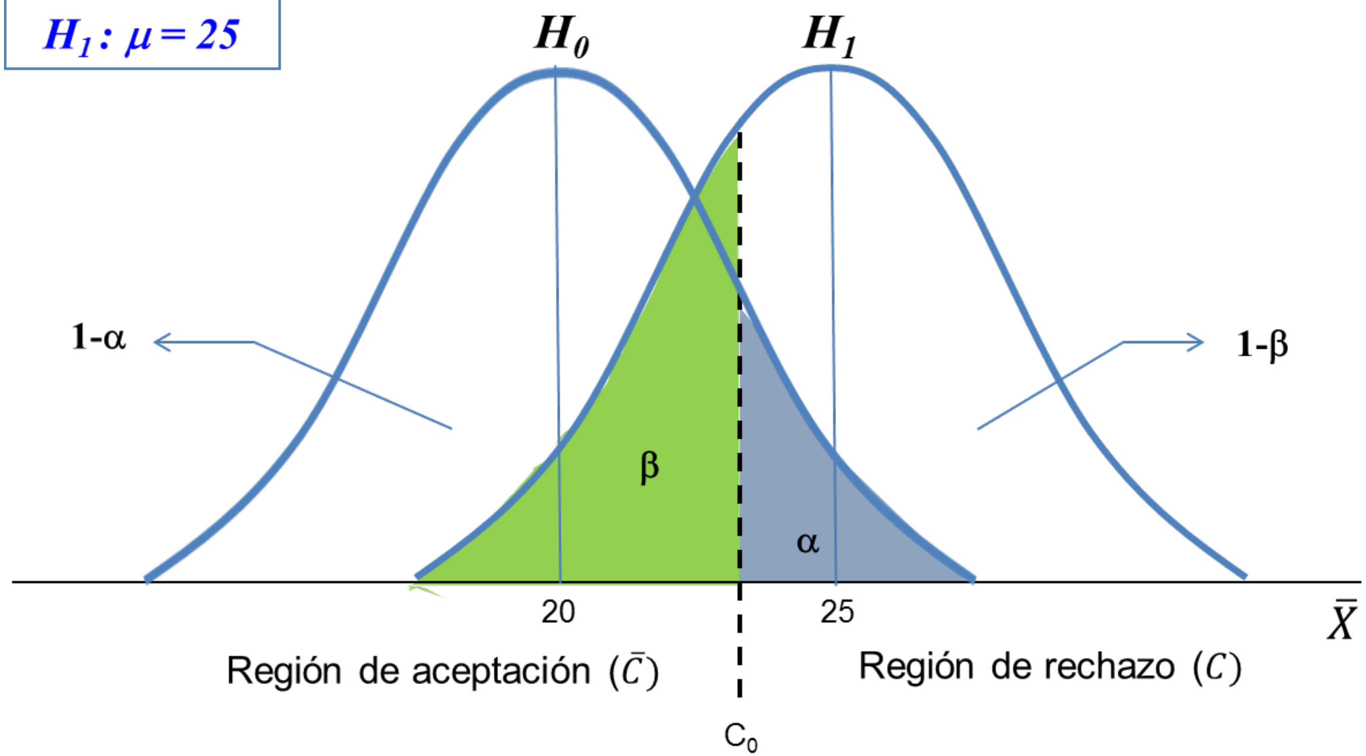
  1-α = P(aceptar H_0/H_0) → nivel de confianza

  k=1-β = P(rechazar H_0/H_1) → potencia del contraste

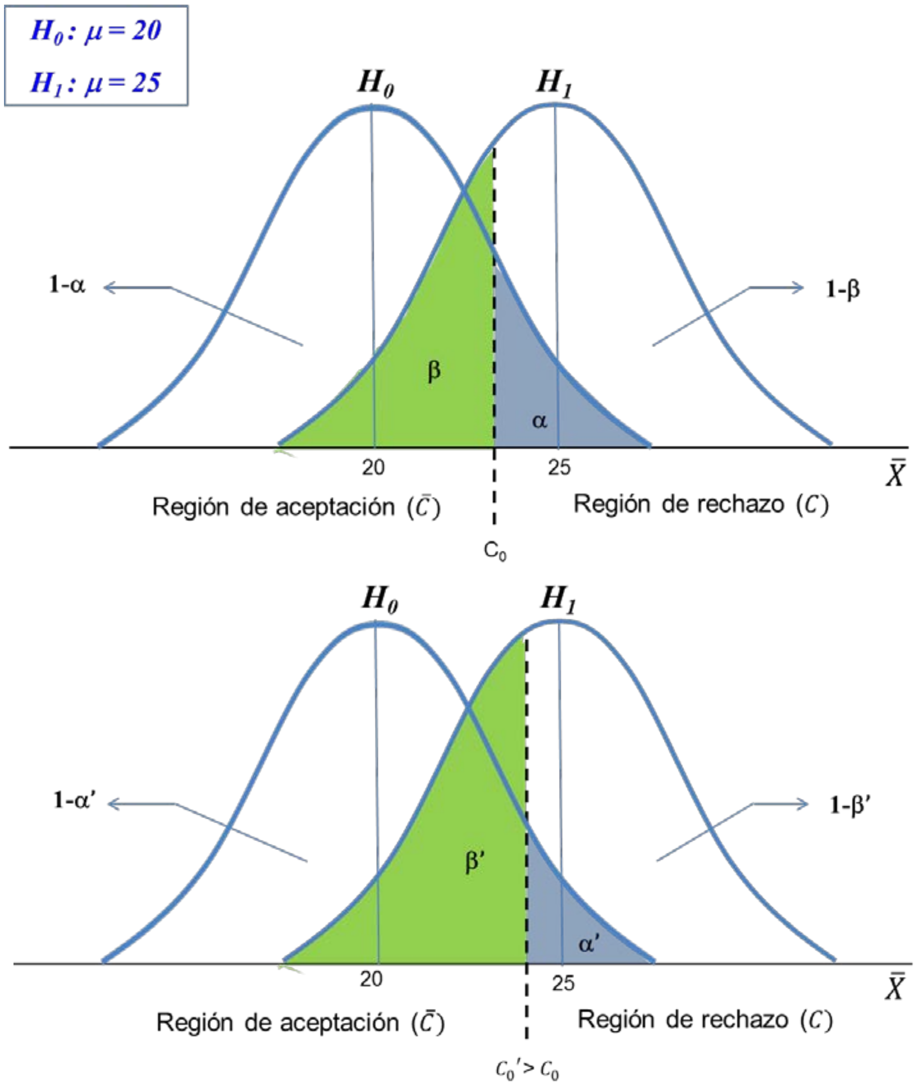
- α y β **no son probabilidades complementarias** (no tienen que sumar 1). La probabilidad contraria de α es 1-α (nivel de significación), y la de β, $1-\beta=k$ (potencia del contraste)

$H_0: \mu = 20$

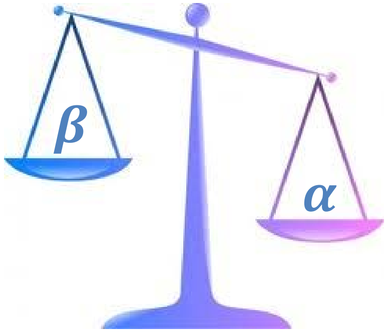
$H_1: \mu = 25$



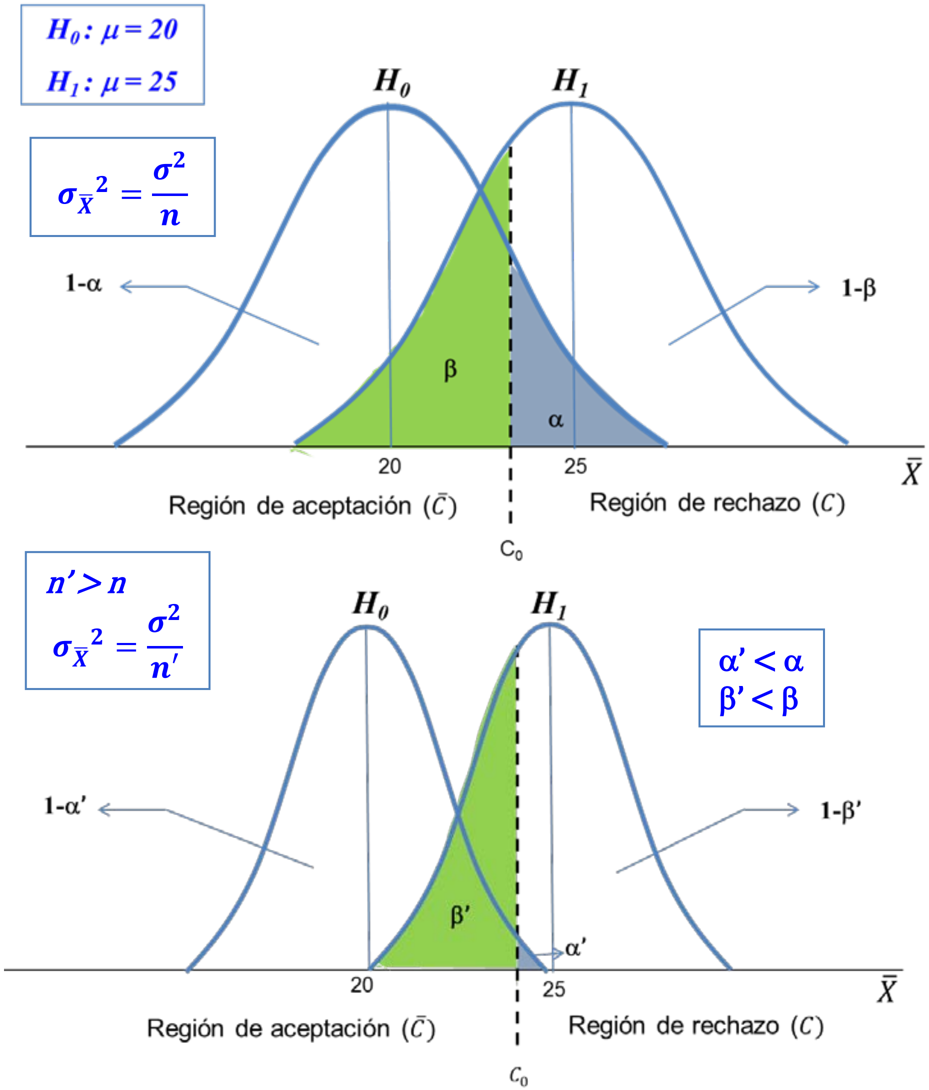
Conceptos básicos



- **α y β no son independientes entre sí:** para un tamaño muestral n, no es posible reducir a la vez ambas probabilidades; si disminuye α aumenta β, y viceversa



Conceptos básicos



- **α y β no son independientes del tamaño de la muestra:**
  - *Si aumentamos el tamaño de la muestra (n) podemos disminuir simultáneamente α y β (ver Gráfico)*
  - *Dado un nivel de* significación α, si n aumenta, disminuye β

Verificación de hipótesis paramétricas

UNIVERSIDAD DE MÁLAGA

- Introducción
- Conceptos básicos
- **Región Crítica Óptima**
- Verificación de la media en poblaciones normales
- Verificación de la varianza en poblaciones normales
- Verificación de la diferencia de medias en poblaciones normales
- Verificación de la proporción
- Verificación de la diferencia de proporciones
- Nivel de significación empírico o p-valor

Región crítica Óptima

- Podemos definir distintas Regiones Críticas para un mismo contraste, pero no todas son iguales en términos de los tamaños de los errores que se pueden cometer

- Diremos que **C** es una **Región Crítica Óptima** de tamaño **α** para verificar $H_0: \theta = \theta_0$ frente a $H_1: \theta = \theta_1$, si para cualquier otro subconjunto A del espacio muestral de igual tamaño (P $[(X_1, X_2, \cdots, X_n) \in A/H_0] = \alpha$), se cumple que:

1) $P[(X_1, X_2, \cdots, X_n) \in C/H_0] = \alpha$

2) $P[(X_1, X_2, \cdots, X_n) \in C/H_1] \geq P[(X_1, X_2, \cdots, X_n) \in A/H_1]$

Región crítica Óptima

- Es decir, entre dos Regiones Críticas de igual tamaño (α), será óptima la que haga que el test tenga mayor potencia (menor β):

$$\boxed{k_C \geq k_A \rightarrow \beta_C \leq \beta_A}$$

- El enfoque de Neyman-Pearson considera que el error de Tipo I (rechazar H_0 cuando es cierta) es más grave que el error de Tipo II (aceptar H_1 cuando es falsa)

El investigador elige de antemano el nivel de significación que está dispuesto a asumir (normalmente, $\alpha \leq 0{,}05$) y elige el test que minimiza β

Región crítica Óptima

UNIVERSIDAD DE MÁLAGA

Procedimientos para obtener la Región Crítica Óptima:

- H_0 y H_1 simples → **Neyman-Pearson**
- H_0 simple y H_1 unilateral → **Test uniformemente más potentes (Neyman-Pearson)**
- Caso general (H_1 bilateral) → **Test de la razón de verosimilitudes**

En el resto del tema se presentan distintos contrastes paramétricos y las ***Regiones Críticas Óptimas*** *que resultan de la aplicación de estos procedimientos (sin entrar en detalles teóricos)*

En resumen, hasta ahora hemos visto que…

UNIVERSIDAD DE MÁLAGA

Contraste de hipótesis

Proceso de decisión entre dos propuestas alternativas (H_0 y H_1), acerca de alguna característica desconocida de una variable, basado en una regla de decisión o test y en la información suministrada por una muestra

Tamaño de los errores y Región Crítica Óptima

El enfoque de **Neyman-Pearson** considera que el **error de Tipo I** (rechazar H_0 cuando es cierta) es más grave que el **error de Tipo II** (aceptar H_1 cuando es falsa).

Entre todas las posibles Regiones Críticas, será **Óptima** aquella que, una vez fijado el riesgo del error tipo I que estamos dispuestos a asumir (nivel de significación $\boldsymbol{\alpha}$, normalmente menor o igual al 5%), minimiza el riesgo del error tipo II ($\boldsymbol{\beta}$)*, es decir, hace máxima la potencia del contraste* ($\boldsymbol{k=1-\beta}$)*.*

En resumen, hasta ahora hemos visto que…

UNIVERSIDAD DE MÁLAGA

Etapas en un contraste de hipótesis

1. Planteamiento: establecer $\mathbf{H_0}$ y $\mathbf{H_1}$
2. Construir la **regla de decisión** (Test, contraste o prueba):
   - Determinar el **estadístico de prueba** (p. ej., la media muestral para verificar hipótesis acerca de la media poblacional) y su distribución de probabilidad
   - Definir la **Región Crítica Óptima** (zona de la distribución del estadístico de prueba que conlleva el rechazo de H_0):

     $\theta_1 > \theta_0$ → cola derecha

     $\theta_1 < \theta_0$ → cola izquierda

     $\theta_1 \neq \theta_0$ → bilateral (dos colas)

   - Fijar el **nivel de significación** (α) y el/los **punto/s crítico/s del test (**C_0**)**
3. Decidir a partir del valor observado del estadístico en la **muestra** y obtener conclusiones

UNIVERSIDAD DE MÁLAGA

- Introducción
- Conceptos básicos
- Región Crítica Óptima
- **Verificación de la media en poblaciones normales**
- Verificación de la varianza en poblaciones normales
- Verificación de la diferencia de medias en poblaciones normales
- Verificación de la proporción
- Verificación de la diferencia de proporciones
- Nivel de significación empírico o p-valor

Verificación de la media (pob. normales)

UNIVERSIDAD
DE MÁLAGA

Verificación de la media

- Varianza poblacional (σ) conocida
- Varianza poblacional (σ) desconocida

Verificación de la media

Regiones críticas para verificar la media en poblaciones normales con varianza conocida

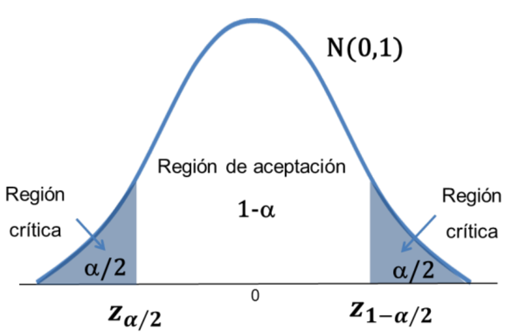
Supuestos necesarios

Normalidad: $\boldsymbol{X \sim N(\mu, \sigma)}$
$\boldsymbol{\sigma}$ conocida
m.a.s. $X_1, X_2, \dots, X_n$

Hipótesis nula	Estadístico de contraste
$H_0 : \mu = \mu_0$	$Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$

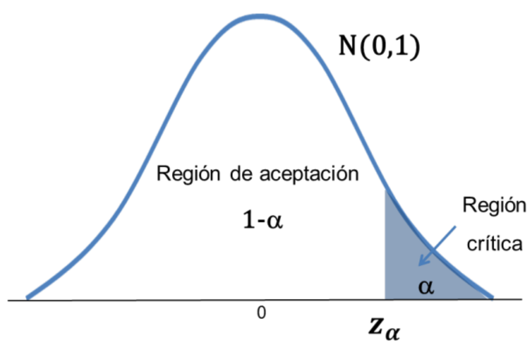
Hipótesis alternativa y Región Crítica Óptima

$H_1 : \mu \neq \mu_0$



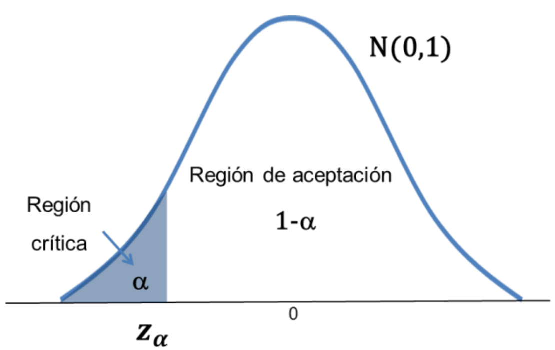
$z_{obs} \geq z_{1-\alpha/2}$ ó $z_{obs} \leq z_{\alpha/2}$

$H_1 : \mu > \mu_0$



$z_{obs} \geq z_{1-\alpha}$

$H_1 : \mu < \mu_0$



$z_{obs} \leq z_{\alpha}$

Verificación de la media

Regiones críticas para verificar la media en poblaciones normales con varianza desconocida

Supuestos necesarios

Normalidad: $\boldsymbol{X \sim N(\mu, \sigma)}$

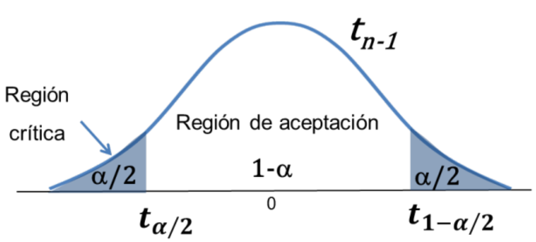
$\boldsymbol{\sigma}$ desconocida

m.a.s. $X_1, X_2, \ldots, X_n$

Hipótesis nula	Estadístico de contraste
$H_0 : \mu = \mu_0$	$\dfrac{\overline{X} - \mu_0}{S / \sqrt{n-1}} = \dfrac{\overline{X} - \mu_0}{\hat{S} / \sqrt{n}} \sim t_{n-1}$

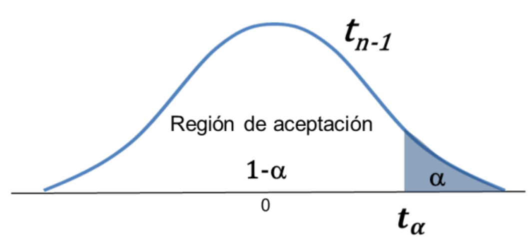
Hipótesis alternativa y Región Crítica Óptima

$H_1 : \mu \neq \mu_0$



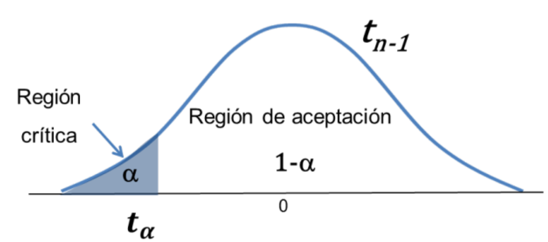
$t_{obs} \geq t_{1-\alpha/2}$ ó $t_{obs} \leq t_{\alpha/2}$

$H_1 : \mu > \mu_0$



$t_{obs} \geq t_{1-\alpha}$

$H_1 : \mu < \mu_0$



$t_{obs} \leq t_{\alpha}$

Verificación de la media en pob. normales

Ejemplo: El dueño de un restaurante en venta afirma que el ingreso medio diario del mismo es de 675 €. Un empresario interesado en adquirirlo duda de esta afirmación y para ver si le interesa realizar la compra selecciona una muestra aleatoria de 25 días en los que obtiene un ingreso medio diario de 600 € y una desviación típica de 125 €. Con un nivel de significación del 5%:

a. Plantee y realice el contraste a partir del cual el empresario tomará su decisión, suponiendo una desviación t poblacional de 150 €. Extraiga conclusiones.

b. Repita el proceso, suponiendo que la varianza poblacional es desconocida

Solución apartado a:

- *Datos:*

X: ventas diarias (€) $X \sim N(\mu, \boldsymbol{\sigma = 150})$

n=25 $\bar{x}$*=600* $\alpha = 0{,}05$

- *Plantear las hipótesis a verificar:*

$H_0: \mu = 675$

$H_1: \mu < 675$

El empresario sólo compra si las ventas no son inferiores a 675 €

- *Determinar el estadístico de prueba (verificación de la media en una población normal con* ***varianza conocida****):*

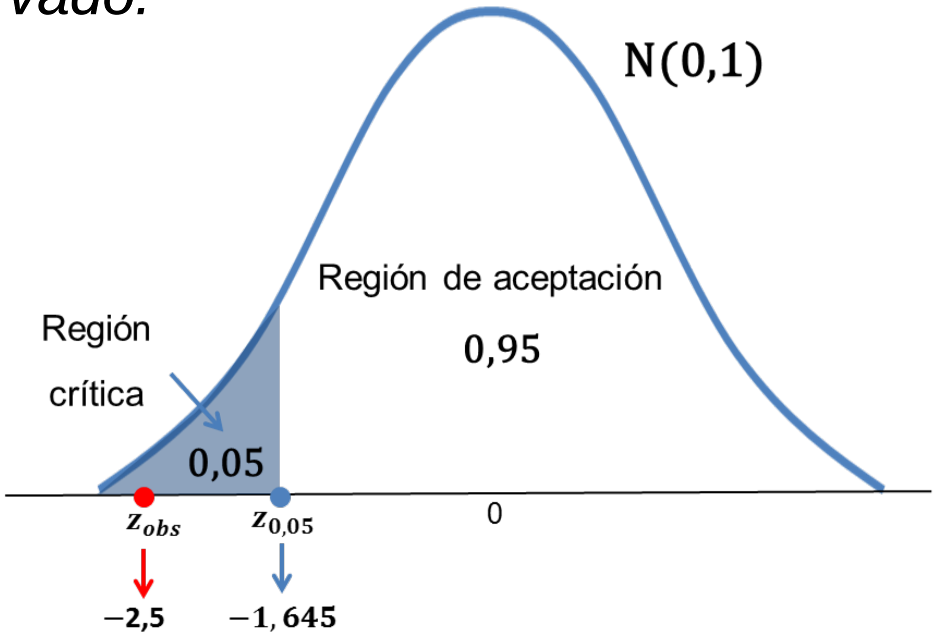
$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$$

Verificación de la media en pob. normales

- *Calcular el valor observado del estadístico de prueba en la muestra:*

$$z_{obs} = \frac{600 - 675}{150/\sqrt{25}} = -2{,}5$$

- *Determinar la región crítica de tamaño α y comparar con el valor observado:*



Verificación de la media en pob. normales

➢ *Tomar la decisión sobre el contraste e interpretarla:*

$$z_{obs} = -2{,}5 < -1{,}645$$

El valor observado del estadístico de prueba ($z_{obs} = -2{,}5$) *es inferior al punto crítico* (-$1{,}645$)*, es decir, pertenece a la región crítica*

Por tanto, con un nivel de significación del 5%:

- *Debemos tomar la decisión de rechazar la hipótesis nula planteada*
- *Se aceptaría que los ingresos medios diarios del establecimiento son inferiores a 675 €*
- *Con este resultado, el empresario no estará interesado en adquirir el restaurante*

Verificación de la media en pob. normales

UNIVERSIDAD DE MÁLAGA

Solución apartado b:

- *Datos:*

X: ventas diarias (€) $X \sim N(\mu, \boldsymbol{\sigma})$

n=25 $\bar{x}$*=600* $\alpha = 0{,}05$

- *Plantear las hipótesis a verificar:*

$H_0 : \mu = 675$

$H_1 : \mu < 675$

⇨ *El empresario sólo compra si las ventas no son inferiores a 675 €*

- *Determinar el estadístico de prueba (verificación de la media en una población normal con **varianza desconocida**):*

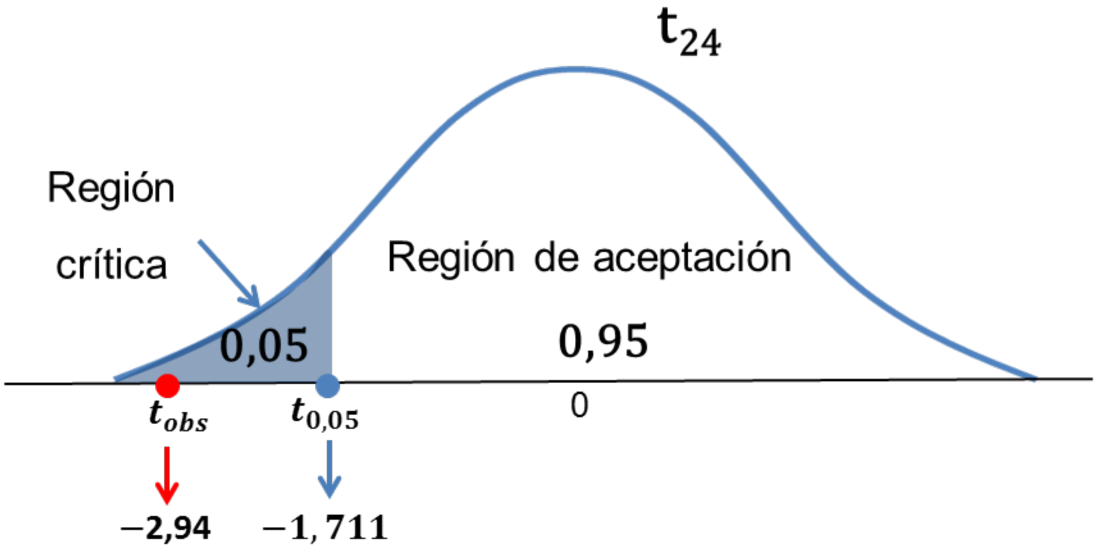
$$\frac{\bar{X} - \mu_0}{S / \sqrt{n-1}} \sim \boldsymbol{t_{n-1}}$$

Verificación de la media en pob. normales

- *Calcular el valor observado del estadístico de prueba en la muestra:*

$$t_{obs} = \frac{600 - 675}{125/\sqrt{24}} = -2{,}94$$

- *Determinar la región crítica de tamaño α y comparar con el valor observado:*



Verificación de la media en pob. normales

- *Tomar la decisión sobre el contraste e interpretarla:*

$$t_{obs} = -2{,}94 < -1{,}711$$

El valor observado del estadístico de prueba $(t_{obs} = -2{,}94)$ *es inferior al punto crítico (*$-1{,}711$*), es decir, pertenece a la región crítica*

Por tanto, con un nivel de significación del 5%:

- *Debemos tomar la decisión de rechazar la hipótesis nula planteada*
- *Se aceptaría que los ingresos medios diarios del establecimiento son inferiores a 675 €*
- *Con este resultado, el empresario no estará interesado en adquirir el restaurante*

Verificación de hipótesis paramétricas

UNIVERSIDAD DE MÁLAGA

- Introducción
- Conceptos básicos
- Región Crítica Óptima
- Verificación de la media en poblaciones normales
- **Verificación de la varianza en poblaciones normales**
- Verificación de la diferencia de medias en poblaciones normales
- Verificación de la proporción
- Verificación de la diferencia de proporciones
- Nivel de significación empírico o p-valor

Verificación de la varianza (pob. normales)

- Verificación de la varianza
  - Media poblacional (μ) conocida
  - Media poblacional (μ) desconocida

Verificación de la varianza

Regiones críticas para verificar la varianza en poblaciones normales con media conocida

Supuestos necesarios

Normalidad: $\boldsymbol{X} \sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{\sigma})$

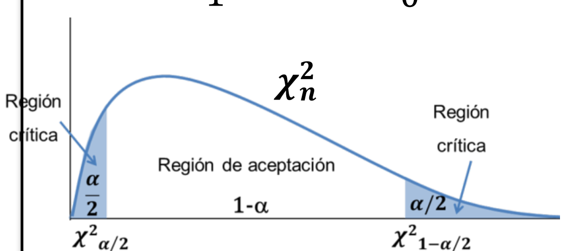
$\boldsymbol{\mu}$ conocida

m.a.s. $X_1, X_2, \dots, X_n$

Hipótesis nula	Estadístico de contraste
$H_0\!: \sigma^2 = \sigma_0^2$	$\dfrac{nS_\mu^2}{\sigma^2} \sim \chi_n^2$ $\qquad S_\mu^2 = \dfrac{\sum (X_i - \mu)^2}{n}$

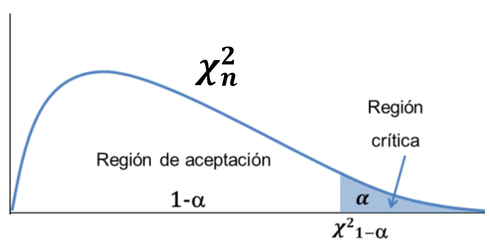
Hipótesis alternativa y Región Crítica Óptima

$$H_1\!: \sigma^2 \neq \sigma_0^2$$



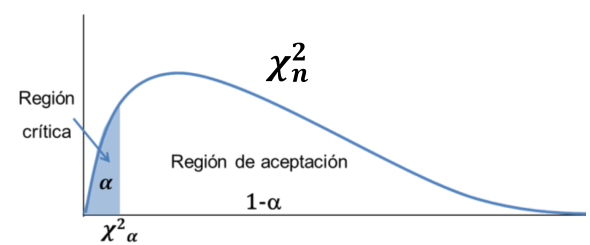
$\chi^2_{obs} \leq \chi^2_{\alpha/2}$ ó $\chi^2_{obs} \geq \chi^2_{1-\alpha/2}$

$$H_1\!: \sigma^2 > \sigma_0^2$$



$\chi^2_{obs} \geq \chi^2_{1-\alpha}$

$$H_1\!: \sigma^2 < \sigma_0^2$$



$\chi^2_{obs} \leq \chi^2_{\alpha}$

Verificación de la varianza

Regiones críticas para verificar la varianza en poblaciones normales con media desconocida

Supuestos necesarios

Normalidad: $\boldsymbol{X \sim N(\mu, \sigma)}$
$\boldsymbol{\mu}$ desconocida
m.a.s. $X_1, X_2, \dots, X_n$

Hipótesis nula

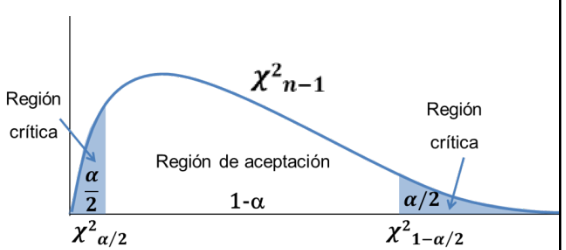
$H_0\colon\ \sigma^2 = \sigma_0^2$

Estadístico de contraste

$$\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1} \qquad S^2 = \frac{\sum (X_i - \overline{X})^2}{n}$$

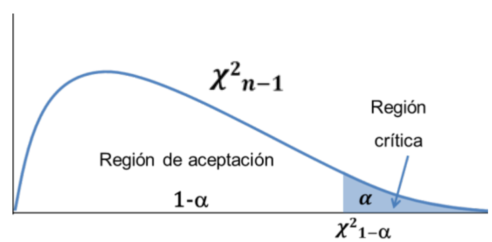
Hipótesis alternativa y Región Crítica Óptima

$H_1\colon\ \sigma^2 \neq \sigma_0^2$



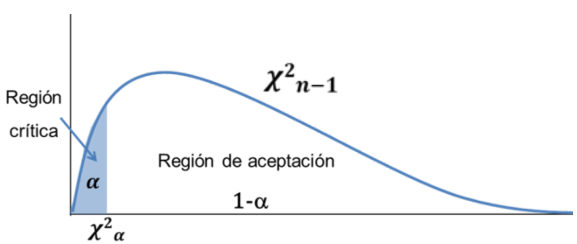
$\chi^2_{obs} \leq \chi^2_{\alpha/2}$ ó $\chi^2_{obs} \geq \chi^2_{1-\alpha/2}$

$H_1\colon\ \sigma^2 > \sigma_0^2$



$\chi^2_{obs} \geq \chi^2_{1-\alpha}$

$H_1\colon\ \sigma^2 < \sigma_0^2$



$\chi^2_{obs} \leq \chi^2_{\alpha}$

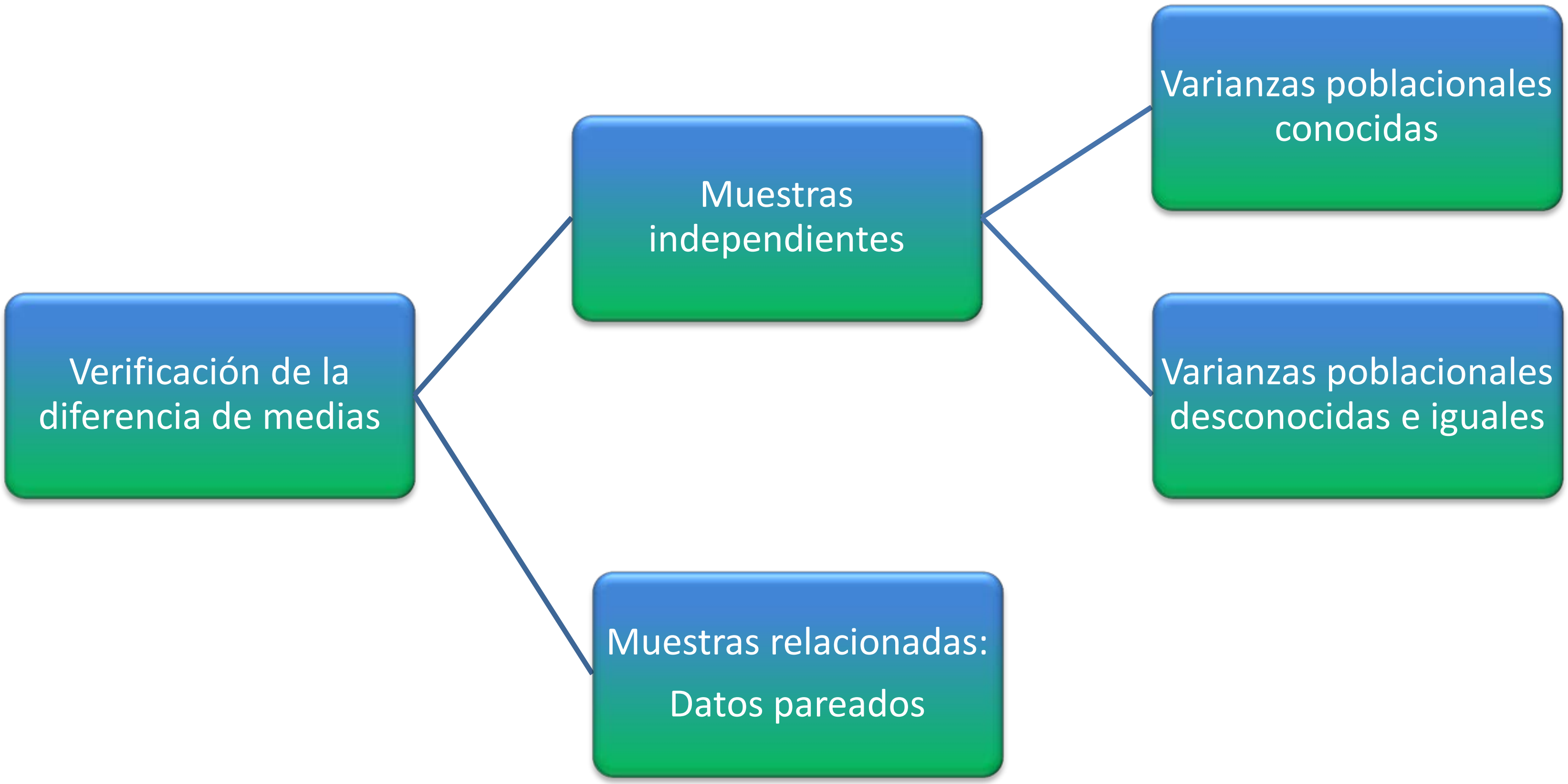
Verificación de hipótesis paramétricas

UNIVERSIDAD DE MÁLAGA

- Introducción
- Conceptos básicos
- Región Crítica Óptima
- Verificación de la media en poblaciones normales
- Verificación de la varianza en poblaciones normales
- **Verificación de la diferencia de medias en poblaciones normales**
- Verificación de la proporción
- Verificación de la diferencia de proporciones
- Nivel de significación empírico o p-valor

Verificación de la diferencia de medias (pob. normales)

UNIVERSIDAD DE MÁLAGA



Verificación de la diferencia de medias

UNIVERSIDAD DE MÁLAGA

Verificación de la diferencia de medias en poblaciones normales.
Muestras independientes y varianzas conocidas

Supuestos necesarios

Normalidad: $\boldsymbol{X \sim N(\mu_x, \sigma_x)}$ $\quad\quad$ $\boldsymbol{Y \sim N(\mu_y, \sigma_y)}$

$\sigma_x^{\,2}$ y $\sigma_y^{\,2}$ conocidas

m.a.s. e independientes: $X_1, X_2, \ldots, X_{n_x}$ $\quad\quad$ $Y_1, Y_2, \ldots, Y_{n_y}$

Hipótesis nula

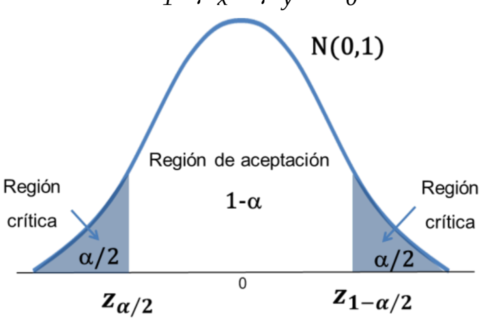
$$H_0 : \mu_x - \mu_y = \delta_0$$

Estadístico de contraste

$$Z = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\dfrac{\sigma_x^2}{n_x} + \dfrac{\sigma_y^2}{n_y}}} \sim N(0,1)$$

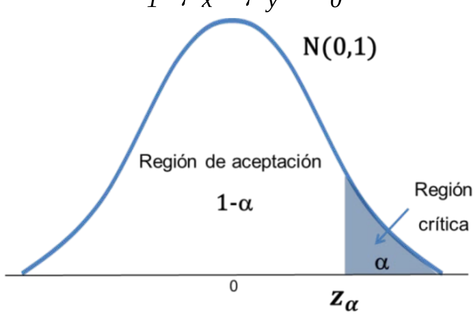
Hipótesis alternativa y Región Crítica Óptima

$$H_1 : \mu_x - \mu_y \neq \delta_0$$



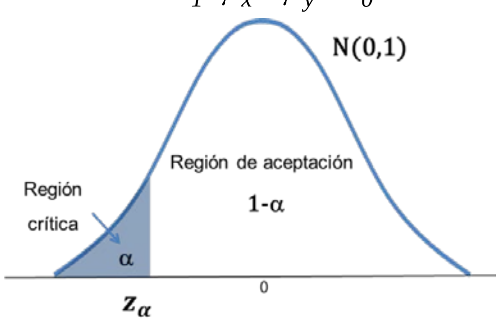
$z_{obs} \geq z_{1-\alpha/2}$ ó $z_{obs} \leq z_{\alpha/2}$

$$H_1 : \mu_x - \mu_y > \delta_0$$



$z_{obs} \geq z_{1-\alpha}$

$$H_1 : \mu_x - \mu_y < \delta_0$$



$z_{obs} \leq z_\alpha$

Verificación de la diferencia de medias

Verificación de la diferencia de medias en poblaciones normales.
Muestras independientes y varianzas desconocidas e iguales $(\sigma_x^2 = \sigma_y^2)$

Supuestos necesarios

Normalidad: $\boldsymbol{X \sim N(\mu_x, \sigma_x)}$ $\boldsymbol{Y \sim N(\mu_y, \sigma_y)}$

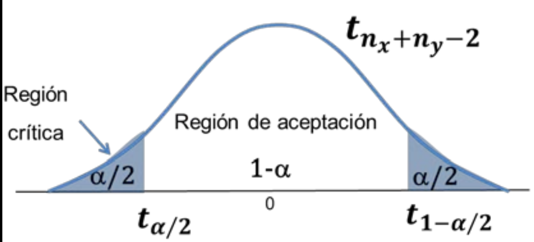
Varianzas poblacionales desconocidas e iguales $(\sigma_x^2 = \sigma_y^2)$

m.a.s. e independientes: $X_1, X_2, \ldots, X_{n_x}$ $Y_1, Y_2, \ldots, Y_{n_y}$

Hipótesis nula	Estadístico de contraste
$H_0 : \mu_x - \mu_y = \delta_0$	$\dfrac{\bar{X} - \bar{Y} - \delta_0}{S_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim t_{n_x + n_y - 2}$ $\quad S_p = \sqrt{\dfrac{n_x S_x^2 + n_y S_y^2}{n_x + n_y - 2}}$

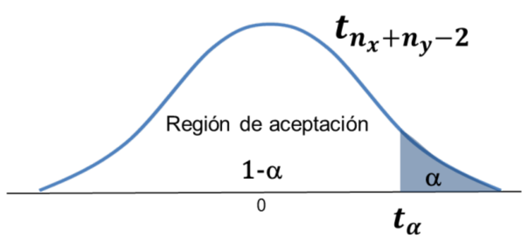
Hipótesis alternativa y Región Crítica Óptima

$H_1 : \mu_x - \mu_y \neq \delta_0$



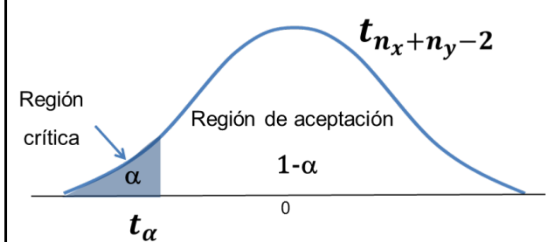
$t_{obs} \geq t_{1-\alpha/2}$ ó $t_{obs} \leq t_{\alpha/2}$

$H_1 : \mu_x - \mu_y > \delta_0$



$t_{obs} \geq t_{1-\alpha}$

$H_1 : \mu_x - \mu_y < \delta_0$



$t_{obs} \leq t_\alpha$

Verificación de la diferencia de medias

UNIVERSIDAD DE MÁLAGA

Regiones críticas para verificar la media en poblaciones normales con datos pareados

Supuestos necesarios

Normalidad: $\boldsymbol{X \sim N(\mu_x, \sigma_x)}$ $\quad$ $\boldsymbol{Y \sim N(\mu_y, \sigma_y)}$

Varianzas poblacionales desconocidas

m.a.s. relacionadas: $(\boldsymbol{X_1}, \boldsymbol{Y_1}), (\boldsymbol{X_2}, \boldsymbol{Y_2}), \ldots, (\boldsymbol{X_n}, \boldsymbol{Y_n})$ **Datos emparejados para cada elemento de la muestra**

Hipótesis nula

$H_0 : \mu_x - \mu_y = \delta_0 \longrightarrow H_0 : \mu_D = \delta_0$

Diferencias para cada par: $D_i = X_i - Y_i \quad i = 1, 2, \ldots, n$

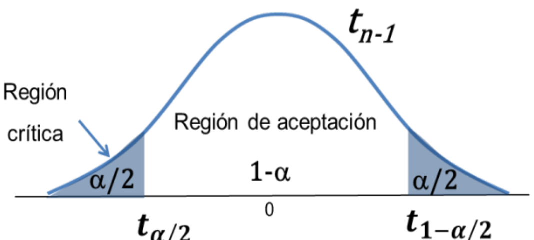
$D_i \sim N(\mu_d, \sigma_d)$

Estadístico de contraste

$$\frac{\overline{D} - \delta_0}{S_D / \sqrt{n-1}} = \frac{\overline{D} - \delta_0}{\hat{S}_D / \sqrt{n}} \sim t_{n-1}$$

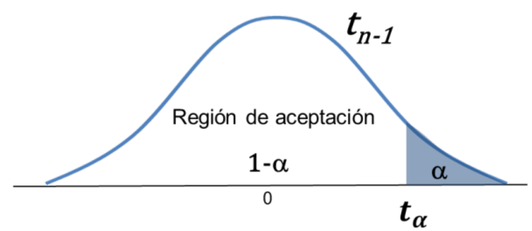
Hipótesis alternativa y Región Crítica Óptima

$H_1 : \mu_x - \mu_y \neq \delta_0 \longrightarrow H_1 : \mu_D \neq \delta_0$



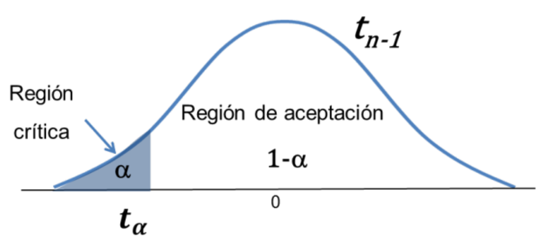
$t_{obs} \geq t_{1-\alpha/2}$ ó $t_{obs} \leq t_{\alpha/2}$

$H_1 : \mu_x - \mu_y > \delta_0 \longrightarrow H_1 : \mu_D > \delta_0$



$t_{obs} \geq t_{1-\alpha}$

$H_1 : \mu_x - \mu_y < \delta_0 \longrightarrow H_1 : \mu_D < \delta_0$



$t_{obs} \leq t_{\alpha}$

UNIVERSIDAD DE MÁLAGA

- Introducción
- Conceptos básicos
- Región Crítica Óptima
- Verificación de la media en poblaciones normales
- Verificación de la varianza en poblaciones normales
- Verificación de la diferencia de medias en poblaciones normales
- **Verificación de la proporción**
- Verificación de la diferencia de proporciones
- Nivel de significación empírico o p-valor

UNIVERSIDAD DE MÁLAGA

Regiones críticas para la verificación de la proporción

Supuestos necesarios

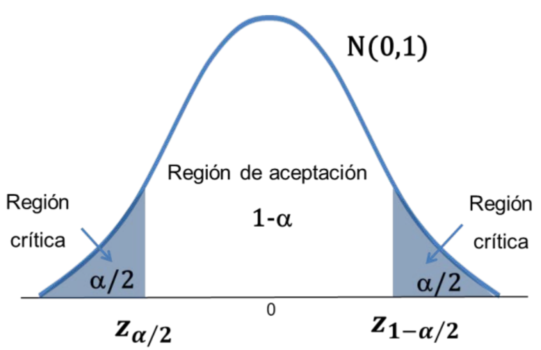
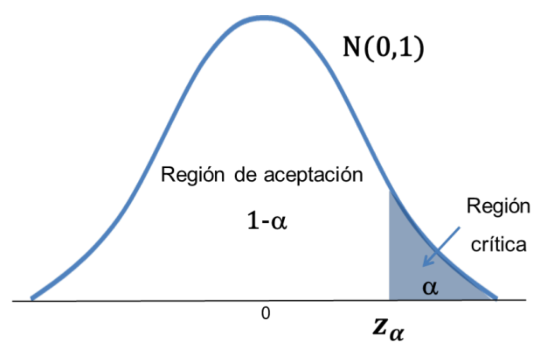
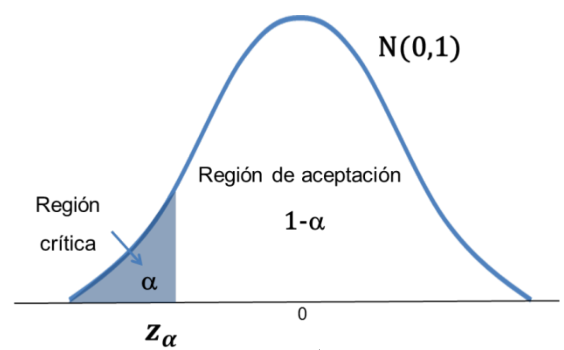
$\boldsymbol{X \sim B(1, p)}$

m.a.s. $X_1, X_2, \dots, X_n$

n grande

Hipótesis nula	Estadístico de contraste
$H_0 : p = p_0$	$Z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}} \xrightarrow[n \to \infty]{} N(0,1)$ $\quad \hat{p} = \dfrac{x}{n} = \dfrac{\text{N}^{\text{o}} \text{ éxitos en la muestra}}{\text{Tamaño de la muestra}}$

Hipótesis alternativa y Región Crítica Óptima

$H_1 : p \neq p_0$	$H_1 : p > p_0$	$H_1 : p < p_0$
		
$z_{obs} \geq z_{1-\alpha/2}$ ó $z_{obs} \leq z_{\alpha/2}$	$z_{obs} \geq z_{1-\alpha}$	$z_{obs} \leq z_{\alpha}$

Verificación de hipótesis paramétricas

UNIVERSIDAD DE MÁLAGA

- Introducción
- Conceptos básicos
- Región Crítica Óptima
- Verificación de la media en poblaciones normales
- Verificación de la varianza en poblaciones normales
- Verificación de la diferencia de medias en poblaciones normales
- Verificación de la proporción
- **Verificación de la diferencia de proporciones**
- Nivel de significación empírico o p-valor

Verificación de hipótesis paramétricas

UNIVERSIDAD DE MÁLAGA

- Introducción
- Conceptos básicos
- Región Crítica Óptima
- Verificación de la media en poblaciones normales
- Verificación de la varianza en poblaciones normales
- Verificación de la diferencia de medias en poblaciones normales
- Verificación de la proporción
- **Verificación de la diferencia de proporciones**
- Nivel de significación empírico o p-valor

Verificación de la diferencia de proporciones

UNIVERSIDAD DE MÁLAGA

Verificación de la diferencia de proporciones

Caso general

$$\begin{pmatrix} p_x - p_y = \delta_0 \\ \delta_0 \neq 0 \end{pmatrix}$$

Caso particular

$$\left(p_x - p_y = 0\right)$$

Verificación de la igualdad de proporciones

Verificación de la diferencia de proporciones

Regiones críticas para la verificación de la diferencia de proporciones. Caso general

Supuestos necesarios

$X \sim B(1, p_x)$ $\quad$ $Y \sim B(1, p_y)$

m.a.s. e independientes: $X_1, X_2, \dots, X_{n_x}$ $\quad$ $Y_1, Y_2, \dots, Y_{n_y}$ $\quad$ n_x y n_y grandes

Hipótesis nula

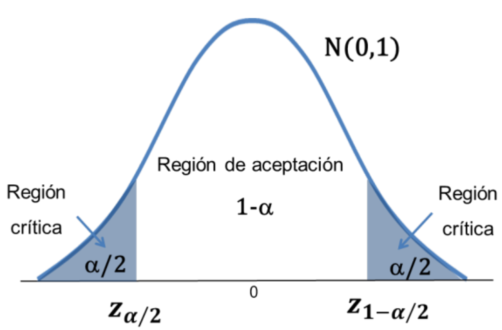
$H_0\colon p_x - p_y = \delta_0$

Estadístico de contraste

$$Z = \frac{\hat{p}_x - \hat{p}_y - \delta_0}{\sqrt{\frac{\hat{p}_x \hat{q}_x}{n_x} + \frac{\hat{p}_y \hat{q}_y}{n_y}}} \xrightarrow[n_x, n_y \to \infty]{} N(0,1) \qquad \hat{p}_x = \frac{x}{n_x} \qquad \hat{p}_y = \frac{y}{n_y}$$

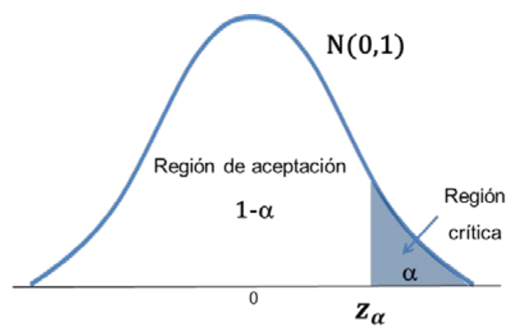
Hipótesis alternativa y Región Crítica Óptima

$H_1\colon p_x - p_y \neq \delta_0$



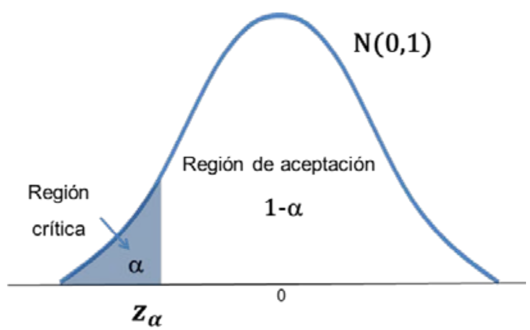
$z_{obs} \geq z_{1-\alpha/2}$ ó $z_{obs} \leq z_{\alpha/2}$

$H_1\colon p_x - p_y > \delta_0$



$z_{obs} \geq z_{1-\alpha}$

$H_1\colon p_x - p_y < \delta_0$



$z_{obs} \leq z_{\alpha}$

Verificación de la diferencia de proporciones

UNIVERSIDAD DE MÁLAGA

Regiones críticas para la verificación de la igualdad de proporciones. Caso particular $(\delta_0 = 0)$

Supuestos necesarios

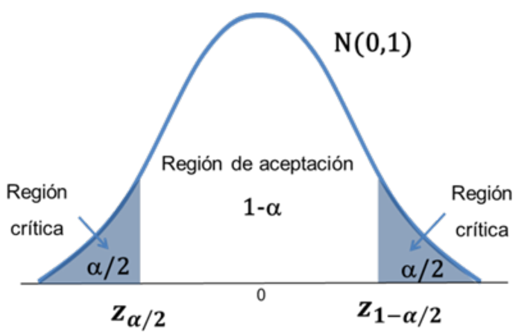
$\boldsymbol{X \sim B(1, p_x)}$ $\quad$ $\boldsymbol{Y \sim B(1, p_y)}$

m.a.s. e independientes: $X_1, X_2, \dots, X_{n_x}$ $\quad$ $Y_1, Y_2, \dots, Y_{n_y}$ $\quad$ n_x y n_y grandes

Hipótesis nula	Estadístico de contraste
$H_0\colon p_x - p_y = 0$	$Z = \dfrac{\hat{p}_x - \hat{p}_y}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}} \xrightarrow[n_x, n_y \to \infty]{} N(0,1) \qquad \hat{p} = \dfrac{n_x \hat{p}_x + n_y \hat{p}_y}{n_x + n_y}$

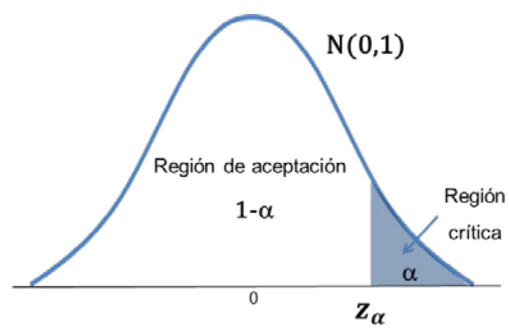
Hipótesis alternativa y Región Crítica Óptima

$H_1\colon p_x - p_y \neq \delta_0$



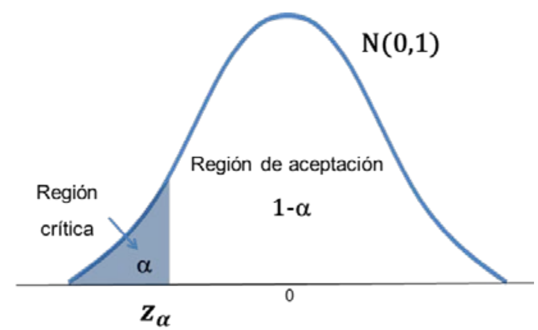
$z_{obs} \leq z_{\alpha/2}$ ó $z_{obs} \geq z_{1-\alpha/2}$

$H_1\colon p_x - p_y > \delta_0$



$z_{obs} \geq z_{1-\alpha}$

$H_1\colon p_x - p_y < \delta_0$



$z_{obs} \leq z_{\alpha}$

Verificación de hipótesis paramétricas

UNIVERSIDAD DE MÁLAGA

- Introducción
- Conceptos básicos
- Región Crítica Óptima
- Verificación de la media en poblaciones normales
- Verificación de la varianza en poblaciones normales
- Verificación de la diferencia de medias en poblaciones normales
- Verificación de la proporción
- Verificación de la diferencia de proporciones
- **Nivel de significación empírico o p-valor**

Nivel de significación empírico (p-valor)

UNIVERSIDAD DE MÁLAGA

- El **p-valor** es el **nivel de significación empírico** del contraste, que se obtiene a partir del valor observado para el estadístico de prueba en la muestra seleccionada
- **P-valor**: probabilidad de obtener, bajo H_0, un valor igual o más extremo al observado (gráficamente: área de la cola correspondiente a la RC a partir del valor observado, o a las dos colas si el contraste es bilateral)

$$0 \leq \text{p-valor} \leq 1$$

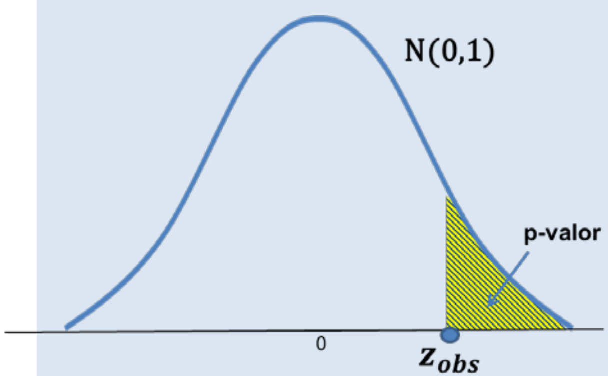
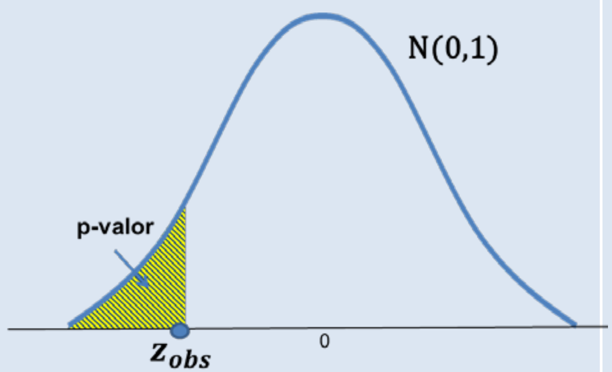
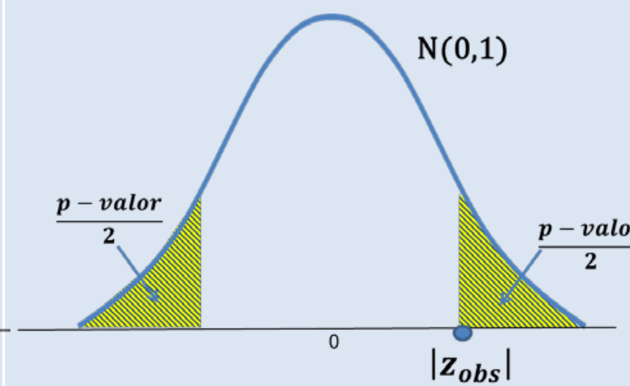
- Permite al investigador decidir por sí mismo el nivel de riesgo α que está dispuesto a asumir, y aplicar la siguiente regla de decisión:

$$\text{si p-valor} < \alpha, \text{ se rechaza } H_0$$

$$\text{si p-valor} > \alpha, \text{ se acepta } H_0$$

Ilustración del cálculo del p-valor

Contraste de la media de una población normal con varianza conocida

$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$
		
$p - valor = pr(Z \geq z_{obs})$	$p - valor = pr(Z \leq z_{obs})$	$p - valor = 2pr(Z \geq \lvert z_{obs} \rvert)$

En los contrastes bilaterales, si la distribución del estadístico de prueba no es simétrica, para calcular el p-valor se multiplica por 2 el área de la cola menor a partir del valor observado

Nivel de significación empírico (p-valor)

UNIVERSIDAD DE MÁLAGA

Ejemplo: Repetir el contraste del ejemplo del restaurante a partir del cálculo del p-valor.

Recordemos datos del apartado a:

X: ventas diarias (€) $X \sim N(\mu, \boldsymbol{\sigma = 150})$

n=25 $\bar{x}$*=600*

$H_0 : \mu = 675$

$H_1 : \mu < 675$

Verificación de la media en una población normal de varianza conocida.

Estadístico de prueba:

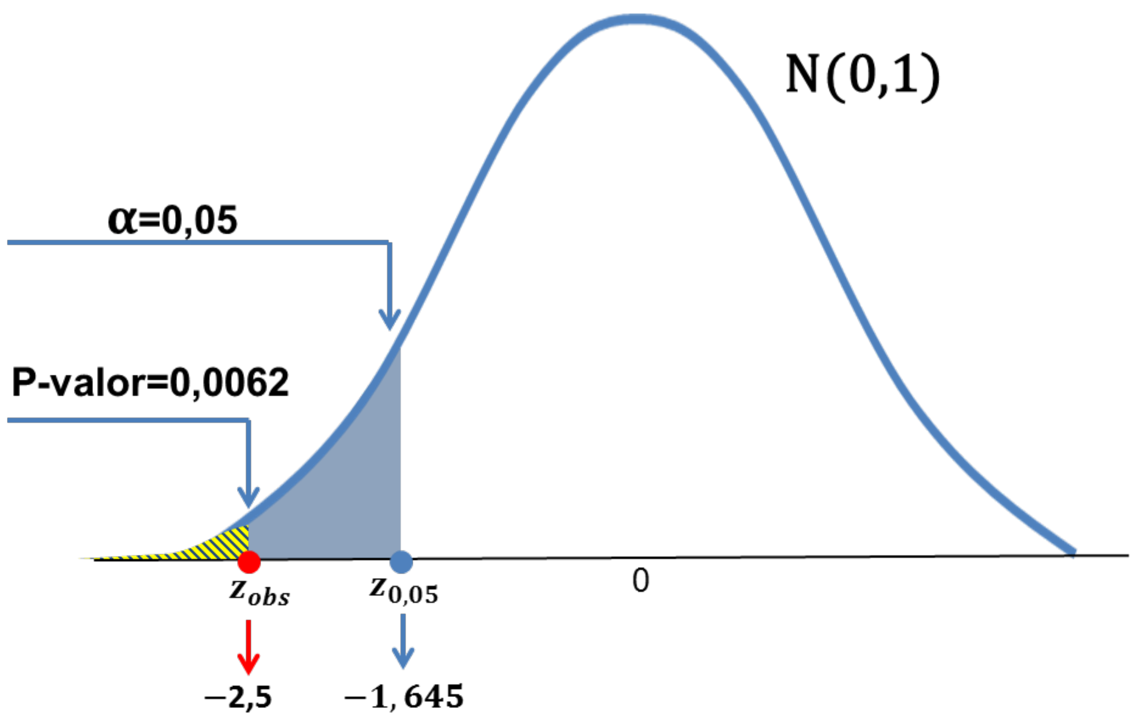
$$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1) \qquad z_{obs} = -2{,}5$$

Nivel de significación empírico (p-valor)

- *Calcular el p-valor a partir del valor observado:*

$$p - valor = pr(Z \leq z_{obs}) = pr(Z \leq -2{,}5) = 0{,}0062$$

- *Comparar el p-valor con* $\boldsymbol{\alpha}$ *:*



Nivel de significación empírico (p-valor)

- *Extraer conclusiones*

Corresponde al investigador fijar el tamaño o riesgo del error de tipo I que está dispuesto a asumir ($\boldsymbol{\alpha}$):

- *En el gráfico, se ha representado la Región Crítica para un $\boldsymbol{\alpha}$ del 5%, de manera que:*

$$\boldsymbol{p - valor = 0,0062 < \alpha = 0,05} \Longrightarrow \boldsymbol{Se\ rechaza\ H_0}$$

- *En realidad, como en este caso el p-valor es muy pequeño, H_0 se rechazaría para los niveles de significación habituales. Sólo para valores de $\boldsymbol{\alpha}$ inferiores a 0,0062 se aceptaría H_0*

Nivel de significación empírico (p-valor)

Apartado b:

X: ventas diarias (€) $X \sim N(\mu, \boldsymbol{\sigma})$

n=25 $\bar{x}$*=600*

$H_0 : \mu = 675$

$H_1 : \mu < 675$

Verificación de la media en una población normal de varianza desconocida.

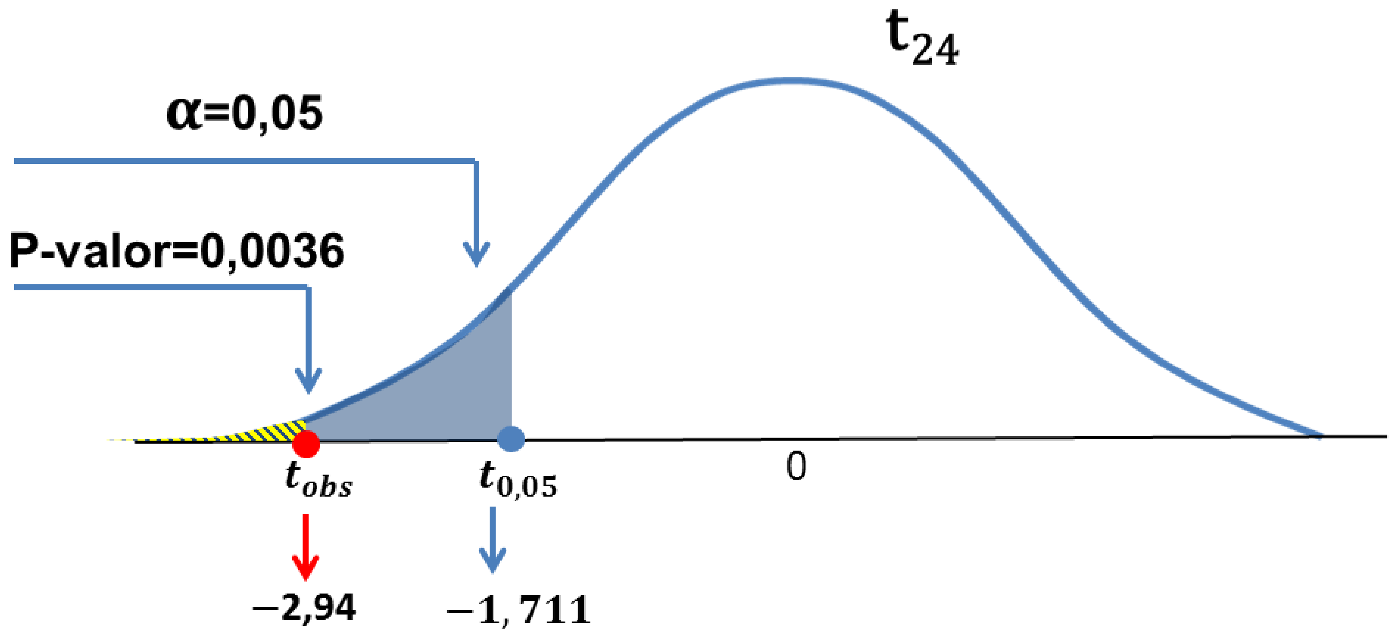
Estadístico de prueba:

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n-1}} \sim \boldsymbol{t_{n-1}} \qquad t_{obs} = -2{,}94$$

$$p - valor = pr(t_{24} \leq t_{obs}) = pr(t_{24} \leq -2{,}94) = 0{,}0036$$

Nivel de significación empírico (p-valor)

En el gráfico, se representa las áreas correspondientes al p-valor y a la Región Crítica del 5%:



$$p - valor = 0,0036 < \alpha = 0,05 \Rightarrow Se\ rechaza\ H_0$$

Verificación de hipótesis paramétricas

Mª Isabel Aguilar, Eugenia Cruces y Bárbara Díaz

UNIVERSIDAD DE MÁLAGA
Departamento de Economía Aplicada (Estadística y Econometría)
Parcialmente financiado a través del PIE13-024 (UMA)