Introducción a la Inferencia Estadistica

Contrastes de Hipótesis

Que es: **hipótesis estadística** es una afirmación respecto a alguna característica de una población.

	Ho verdadera	Ho falsa
DECISIÓN: Mantener Ho	Decisión correcta	Decisión incorrecta Error de tipo II
DECISIÓN: Rechazar Ho	Decisión incorrecta Error de tipo I	Decisión correcta

Ho : **Hipótesis nula**
H1 : **Hipótesis alternativa**

Errores que se pueden cometer

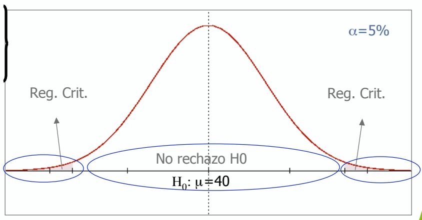
Pueden ser **unilaterales** o **bilaterales**

Conclusiones a partir de una muestra aleatoria y significativa, permite aceptar o rechazar la hipótesis nula

Introducción a la Inferencia Estadistica

Contrastes de Hipótesis

Ho : $\mu = \mu_o$
H1 : $\mu \neq \mu_o$

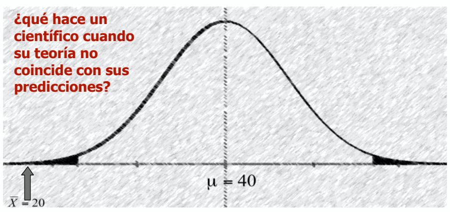


Contraste bilateral

Introducción a la Inferencia Estadistica

Contrastes de Hipótesis

Si supongo que H_0 es cierta...

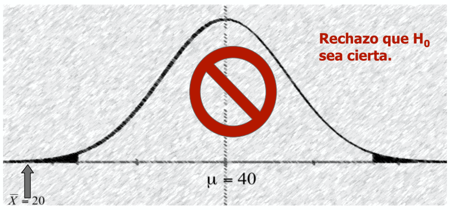


... el resultado del experimento sería improbable.
Sin embargo **ocurrió**.

Introducción a la Inferencia Estadistica

Contrastes de Hipótesis

Si supongo que H_0 es cierta…



… el resultado del experimento sería improbable.

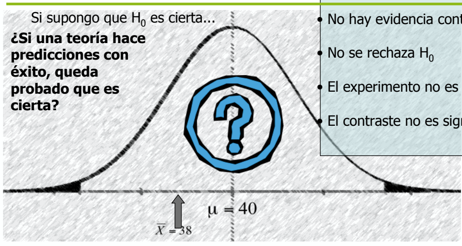
Sin embargo **ocurrió**.

Introducción a la Inferencia Estadistica

Contrastes de Hipótesis

Si supongo que H_0 es cierta…

¿Si una teoría hace predicciones con éxito, queda probado que es cierta?



- No hay evidencia contra H_0
- No se rechaza H_0
- El experimento no es concluyente
- El contraste no es significativo

… el resultado del experimento es coherente.

Introducción a la Inferencia Estadistica

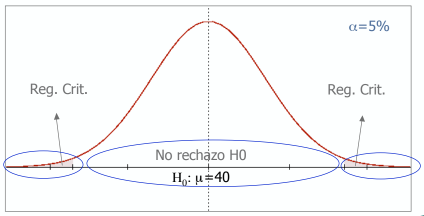
Contrastes de Hipótesis

Región crítica

- Es conocida antes de realizar el experimento: resultados experimentales que refutarían H_0

Nivel de significación: α

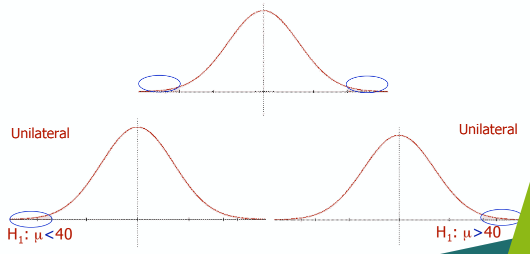
- Número pequeño: 1% , 5%
- Fijado *apriori* por el investigador
- Probabilidad de rechazar H_0 cuando es cierta



Introducción a la Inferencia Estadistica

Contrastes de Hipótesis

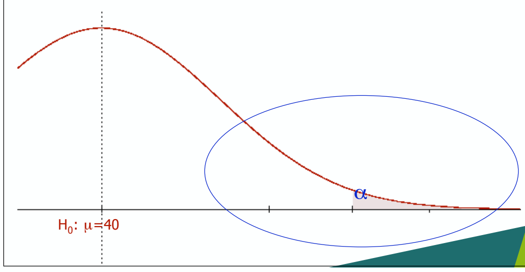
La posición de la región crítica depende de la hipótesis alternativa



Introducción a la Inferencia Estadistica

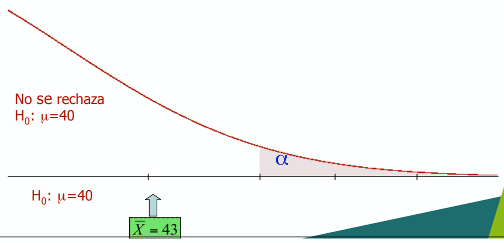
Contrastes de Hipótesis

Significación: p-valor



Introducción a la Inferencia Estadistica

Contrastes de Hipótesis



Introducción a la Inferencia Estadistica

Contrastes de Hipótesis

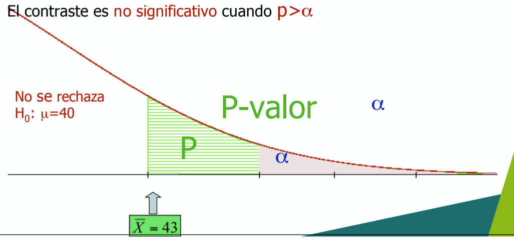
Significacion. P-valor

Probabilidad de tener una muestra que discrepe aún más que la nuestra de H_0.

Probabilidad de obtener una muestra "más extraña" que la obtenida.

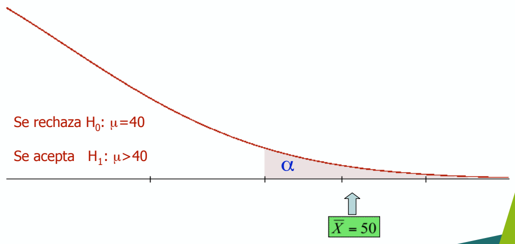
P-valor es conocido después de realizar el experimento aleatorio

El contraste es no significativo cuando $p>\alpha$



Introducción a la Inferencia Estadistica

Contrastes de Hipótesis

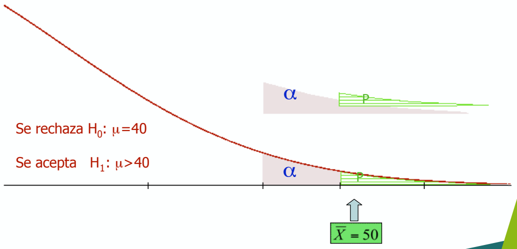


Introducción a la Inferencia Estadistica

Contrastes de Hipótesis

El contraste es estadísticamente significativo cuando $p<\alpha$

Es decir, si el resultado experimental discrepa más de "lo tolerado" a priori.



Introducción a la Inferencia Estadistica

Contrastes de Hipótesis

- Sobre α
  - Número pequeño, elegido *a priori antes de* diseñar el experimento
  - Conocido α sabemos todo sobre la región crítica

- Sobre p
  - Es conocido tras realizar el experimento
  - Conocido p sabemos todo sobre el resultado del experimento

- Sobre el criterio de rechazo
  - Contraste significativo = p menor que α

Introducción a la Inferencia Estadistica

Contrastes de Hipótesis

- Estamos estudiando el efecto del estrés sobre la presión arterial. Nuestra hipótesis es que la presión sistólica media en varones jóvenes estresados es mayor que 18 cm de Hg. Estudiamos una muestra de 36 sujetos:

  $\bar{X}$ = 18,5 S = 3,6

- Plantear Contraste
- Estadístico
- Zona de Rechazo
- P-valor
- Conclusiones
- β

$t(35)_{0,05}$=1,69
T=0,833
No esta en Región Crítica
No es > 1,69
P-valor para T=0,833, y para 35 g.l. es aproximadamente 0,20

NO Se rechaza H_0: μ<=18

H_1: μ>18

Introducción a la Inferencia Estadistica

Contrastes de Hipótesis

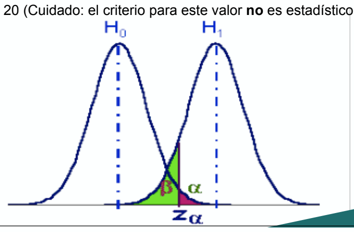
El contraste es estadísticamente significativo cuando $p<\alpha$

Pero que pasa cuando NO Rechazo?

El error cometido es β

Para calcularla se debe concretar H_1

μ_1= 20 (Cuidado: el criterio para este valor **no** es estadístico)



Introducción a la Inferencia Estadistica

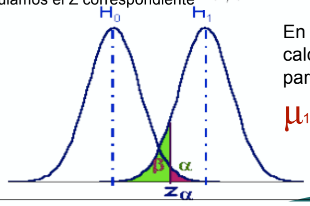
Contrastes de Hipótesis

Cuando Acepto Ho: $\frac{\bar{X}-18}{3,6/\sqrt{36}} \leq 1,69 \Rightarrow \bar{X}-18 \leq 1,01 \Rightarrow \bar{X} \leq 19,01$

El error β es

$P(Z<z_\alpha)$ tomando $\mu_1 = 20$ $\quad \frac{\bar{X}-20}{3,6/\sqrt{36}} \Rightarrow z = \frac{19,01-20}{3,6/\sqrt{36}} = 1,65 \Rightarrow \beta = 0,05$

Calculamos el Z correspondiente

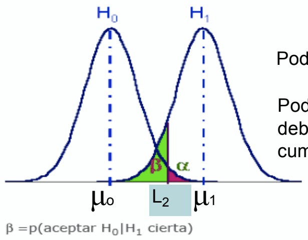


En este caso hemos calculado β para un **n** dado y para una μ_1

Introducción a la Inferencia Estadistica

Contrastes de Hipótesis

Calculo del Tamaño Muestral
Se obtiene a partir de L_2



Podemos fijar **n** y calcular β

ó

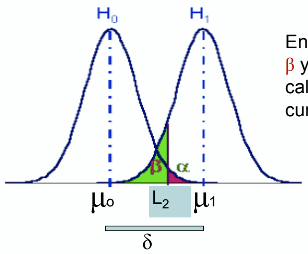
Podemos fijar β y entonces debemos calcular **n** para cumplir con ese error

Introducción a la Inferencia Estadistica

Contrastes de Hipótesis

Calculo del Tamaño Muestral
Se obtiene a partir de L_2

$$n = (z_\alpha + z_\beta)^2 \, (\sigma/\delta)^2$$

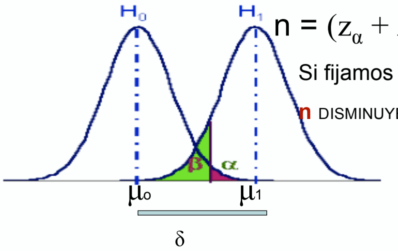


En este caso hemos fijado β y entonces debemos calcular el **n** para cumplir con ese error

Introducción a la Inferencia Estadistica

Contrastes de Hipótesis

Comportamiento de β, δ y el tamaño muestral



$$n = (z_\alpha + z_\beta)^2\ (\sigma/\delta)^2$$

Si fijamos β

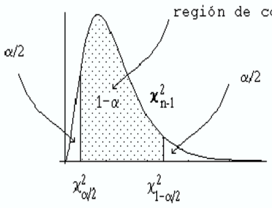
n DISMINUYE cuando aumento δ

Introducción a la Inferencia Estadistica

Intervalo de Confianza

- Intervalo de Confianza para la Varianza

$$\mathcal{P}\left[\chi^2_{n-1,\alpha/2} \le \chi^2_{n-1} \le \chi^2_{n-1,1-\alpha/2}\right] = 1-\alpha$$



$$\chi^2_{n-1,\alpha/2} \le \frac{(n-1)\hat{S}^2}{\sigma^2} \le \chi^2_{n-1,1-\alpha/2}$$

$$\frac{(n-1)\hat{S}^2}{\chi^2_{n-1,1-\alpha/2}} \le \sigma^2 \le \frac{(n-1)\hat{S}^2}{\chi^2_{n-1,\alpha/2}}$$

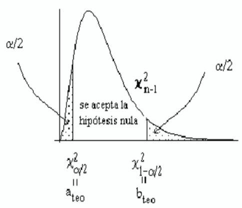
$$\sigma^2 \in \left[\frac{(n-1)\hat{S}^2}{\chi^2_{n-1,1-\alpha/2}}\ ,\ \frac{(n-1)\hat{S}^2}{\chi^2_{n-1,\alpha/2}}\right]$$

Introducción a la Inferencia Estadistica

Contrastes de Hipótesis para Varianza

- Plantear Contraste

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \ne \sigma_0^2 \end{cases}$$



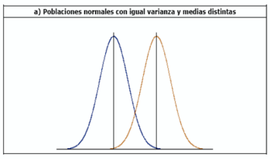
Estadístico

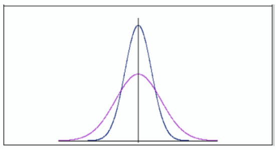
$$\chi^2_{exp} = (n-1)\cdot\frac{\hat{S}^2}{\sigma_0^2}$$

si $a_{teo} \le \chi^2_{exp} \le b_{teo}$ $\Longrightarrow$ no rechazamos H_0;

Introducción a la Inferencia Estadistica

Contrastes para dos Poblaciones





Introducción a la Inferencia Estadistica

Contrastes para dos Poblaciones Independientes

Varianzas Conocidas

$$\left.\begin{array}{l} H_0: \mu_1 - \mu_2 = \Delta_0 \\ H_A: \mu_1 - \mu_2 \neq \Delta_0 \end{array}\right\}$$

$$\left.\begin{array}{l} H_0: \mu_1 - \mu_2 = \Delta_0 \\ H_A: \mu_1 - \mu_2 > \Delta_0 \end{array}\right\}$$

$$Z_{expt} = \frac{(\bar{x} - \bar{y}) - \Delta_0}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

Bilateral

$$P\text{-valor} = 2\,P(z > |z_{expt}|)$$

Unilateral

$$P\text{-valor} = P(z > z_{expt})$$

Introducción a la Inferencia Estadistica

Contrastes para dos Poblaciones Independientes

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$ (con σ_1, σ_2 desconocidas)

región crítica $R = \{|T| > t_{m;\,\alpha/2}\} = \left\{\left|\frac{\overline{X} - \overline{Y}}{S^*}\right| > t_{m,\,\alpha/2}\right\}$

$$\begin{cases} H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

región crítica $R = \{ F \notin (F_{n_1-1, n_2-1; 1-\alpha/2}, F_{n_1-1, n_2-1; \alpha/2})\}$

Varianzas Iguales

$$S^* = \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, S_p$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$m = n_1 + n_2 - 2$$

Varianzas Distintas

$$m = \left[\frac{(S_1^2/n_1 + S_2^2/n_2)^2}{\frac{1}{n_1 - 1}(S_1^2/n_1)^2 + \frac{1}{n_2 - 1}(S_2^2/n_2)^2}\right]$$

$$S^* = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Introducción a la Inferencia Estadistica

Contrastes para dos Poblaciones Independientes

Intervalo de confianza para diferencia de medias
Varianzas Conocidas

$$(\bar{x} - \bar{y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}}$$

Introducción a la Inferencia Estadistica

Contrastes para dos Poblaciones Independientes

Problema.
Se quiere comprobar si los datos de número de resfriados durante el periodo de exámenes por estudiante influyen tomando vitamina C. Los sig. datos corresponden a muestras a las que se les ha suministrado placebo (*x*) y Acido Ascorbico (*y*).

***x* = 2,2**	***y = 1,9***
Sx = 0,12	**Sy= *0,10***
***Nx*= 155**	**Ny = *208***

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases} \quad \text{(con } \sigma_1, \sigma_2 \text{ desconocidas)}$$

Introducción a la Inferencia Estadistica

Contrastes para dos Poblaciones Independientes

Placebo (*x*) y Acido Ascorbico (*y*).

***x* = 2,2**	***y = 1,9***
Sx = 0,12	**Sy= *0,10***
***Nx*= 155**	**Ny = *208***

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases} \quad \text{(con } \sigma_1, \sigma_2 \text{ desconocidas)}$$

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases} \quad \text{región crítica R} = \{ F \notin (F_{n_1-1, n_2-1; 1-\alpha/2}, F_{n_1-1, n_2-1; \alpha/2}) \}$$

$$\frac{\sigma_x^2}{\sigma_y^2} < F_{Nnum-1,\, Nden-1,\, 0.025}$$

$$1{,}2 < 1{,}3637$$ **Varianzas Iguales**

Introducción a la Inferencia Estadistica

Contrastes para dos Poblaciones Independientes

Placebo (x) y Acido Ascorbico (y).

x = 2,2	***y = 1,9***
Sx = 0,12	**Sy*= 0,10***
***Nx*= 155**	**N*y = 208***

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$ (con σ_1, σ_2 desconocidas) (pero iguales)

$$S^* = \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, S_p \qquad t_{obs} = \frac{(\bar{y}_1 - \bar{y}_2) - \Delta_0}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$m = n_1 + n_2 - 2$$

Introducción a la Inferencia Estadistica

Contrastes para dos Poblaciones Independientes

Problema. 2
Los sig. datos corresponden a muestras de dos ciudades
En las que se ha observado el nivel de contaminación de plomo en el agua corriente. Verificar si hay diferencias entre ambas ciudades.

	x	***y***
Mean	96.92	96.31
Std Dev	7.51	8.06
Sample Size	15	16

$$\left.\begin{array}{l} H_0\text{: } \mu_1 - \mu_2 = 0 \\ H_A\text{: } \mu_1 - \mu_2 \neq 0 \end{array}\right\} \quad (\alpha = 0.05)$$

Introducción a la Inferencia Estadistica

Contrastes para dos Poblaciones Independientes

$$S_p^2 = \frac{(15-1)(7.51)^2 + (16-1)(8.06)^2}{15 + 16 - 2} = 60.83$$

$$TS : t_{obs} = \frac{96.92 - 96.31}{\sqrt{60.83\left(\frac{1}{15} + \frac{1}{16}\right)}} = \frac{0.61}{2.80} = 0.22$$

$$RR : |t_{obs}| \geq t_{.025,29} = 2.045$$

$$95\%\, CI : 0.61 \pm 2.045(2.80) \equiv 0.61 \pm 5.73 \equiv (-5.12, 6.34)$$

Introducción a la Inferencia Estadistica

Contrastes para dos Poblaciones Dependientes

$$\left.\begin{array}{l} H_0{:}\mu_D=\Delta_0 \\ H_A{:}\ \mu_D > \Delta_0 \end{array}\right\}$$

$$t_{obs} = \frac{\overline{d}}{s_d / \sqrt{n}}$$

1) Definimos la variable D como:

D = {Antes - Despues} ó
D = {Despues - Antes} ó
D = { X - Y} ó
D = { Y - X}

2) Proponemos el contraste

3) Calculamos el estadístico en función de la definición de D y del contraste propuesto, siendo D <u>una variable Normal</u>

4) Obtenemos las conclusiones y aportamos el p-valor