Tema 7 Inferencia con muestras grandes

Carlos Montes – uc3m

1. Concepto de intervalo de confianza

Intervalo de confianza para el parámetro θ con nivel de confianza (1- α) es el intervalo $\left[\theta_1(X),\theta_2(X)\right]$ tal que:

$$P[\theta_1(X) \le \theta \le \theta_2(X)] = 1 - \alpha$$

La probabilidad de que el intervalo aleatorio (θ_1,θ_2) contenga al verdadero valor del parámetro es 1- α

- 1. Concepto de intervalo de confianza
- 2. Intervalo de confianza para la media
 - 2.1. Varianza poblacional conocida
 - 2.2. Varianza poblacional desconocida
- 3. T de Student
- 4. Otros intervalos de confianza
- 4.1. Para la varianza de poblaciones normales
 - 4.2. Para una proporción
 - 4.3. Para la λ de Poisson
- 5. Determinación del tamaño muestral
- 6. Contraste de hipótesis
 - 6.1. Concepto
 - 6.2. Método
 - 6.3. p-valor
- 7. Contraste para la media
- 8. Contrastes e intervalos

1. Concepto de intervalo de confianza



De cada 100 intervalos construidos a partir de 100 muestras tendrán el $(1-\alpha)100\%$ contendrán el verdadero valor del parámetro

1. Concepto de intervalo de confianza

Dado un α , o nivel de significación se trata de encontrar un intervalo centrado en el parámetro que contenga su verdadero valor el $(1-\alpha)100\%$ de las veces.

Carlos Montes - uc3m

2.1. Intervalo de confianza para la media con σ conocida

Por el teorema del límite central sabemos que:

$$\overline{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Tipificando:

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \to N(0,1) = Z$$

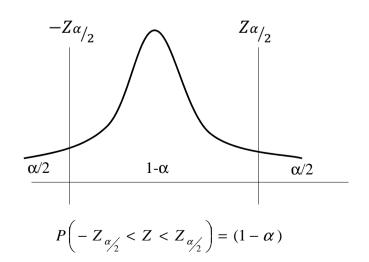
2.1. Intervalo de confianza para la media con σ conocida

Definimos $Z_{\alpha\!\!/_{\!\!2}}$ como el valor de Z tal que:

$$P(Z > Z_{\alpha/2}) = \frac{\alpha}{2}$$

$$P(Z < -Z_{\alpha/2}) = \frac{\alpha}{2}$$

2.1. Intervalo de confianza para la media con σ conocida



2.1. Intervalo de confianza para la media con σ conocida

$$P\left(-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}\right) = (1 - \alpha)$$

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \to N(0, 1) = Z$$

$$P\left(-Z_{\frac{\alpha}{2}} < \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} < Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

Carlos Montes – uc3m

Una muestra aleatoria extraída de una población con σ^2 = 100, y de n = 144 observaciones tiene una media muestral de 160.

Se pide:

- a) Calcular un intervalo de confianza del 95% para la media poblacional μ .
- b) Calcular un intervalo de confianza del 90% para la media poblacional μ .

2.1. Intervalo de confianza para la media con σ conocida

$$P(-Z_{\alpha/2} \le \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le Z_{\alpha/2}) = 1 - \alpha$$

$$P(\overline{X} - Z_{\alpha/2}) = 1 - \alpha$$

$$P(\overline{X} - Z_{\alpha/2}) = 1 - \alpha$$

$$\mu \in \left(\overline{X} \pm Z\alpha_{/2} \cdot \sigma_{/\sqrt{n}}\right)$$

$$\mu \in \left(\bar{X} \pm Z\alpha_{/2} \cdot \sigma / \sqrt{n}\right)$$

$$\mu \in \left(160 \pm Z\alpha_{/2} \cdot 10 / 12\right)$$

a)
$$1 - \alpha = 0.95$$
 $\alpha = 0.05$ $P\left(Z > Z\alpha_{/2}\right) = \frac{0.05}{2} = 0.025$ $P\left(Z < Z\alpha_{/2}\right) = 1 - 0.025 = 0.975$

P(Z <z)< th=""><th>0,00</th><th>0,01</th><th>0,02</th><th>0,03</th><th>0,04</th><th>0,05</th><th>0,06</th><th>0,07</th><th>0,08</th><th>0,09</th></z)<>	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,500	0,504	0,508	0,512	0,516	0,520	0524	0,528	0,532	0,5
0,1	0,540	0,544	0,548	0,552	0,556	0,560	0 564	0,567	0,571	0,5
0,2	0,579	0,583	0,587	0,591	0,595	0,599	0 603	0,606	0,610	0,6
0,3	0,618	0,622	0,626	0,629	0,633	0,637	0 641	0,644	0,648	0,6
0,4	0,655	0,659	0,663	0,666	0,670	0,674	0 677	0,681	0,684	0,6
0,5	0,691	0,695	0,698	0,702	0,705	0,709	0 712	0,716	0,719	0,7
0,6	0,726	0,729	0,732	0,736	0,739	0,742	0 745	0,749	0,752	0,7
0,7	0,758	0,761	0,764	0,767	0,770	0,773	0 776	0,779	0,782	0,7
0,8	0,788	0,791	0,794	0,797	0,800	0,802	0 805	0,808	0,811	0,8
0,9	0,816	0,819	0,821	0,824	0,826	0,829	0831	0,834	0,836	0,8
1	0,841	0,844	0,846	0,848	0,851	0,853	0 855	0,858	0,860	0,8
1,1	0,864	0,867	0,869	0,871	0,873	0,875	0 877	0,879	0,881	0,8
1,2	0,885	0,887	0,889	0,891	0,893	0,894	0 896	0,898	0,900	0,9
1,3	0,903	0,905	0,907	0,908	0,910	0,911	0 913	0,915	0,916	0,9
1,4	0,919	0,921	0,922	0,924	0,925	0,926	0 928	0,929	0,931	0,9
1,5	0,933	0,934	0,936	0,937	0,938	0,939	0 941	0,942	0,943	0,9
1,6	0,945	0,946	0,947	0,948	0,949	0,951	0 952	0,953	0,954	0,9
1,7	0,955	0,956	0,957	0,958	0,959	0,960	0 961	0,962	0,962	0,9
1,8	0,964	0,965	0,966	0,966	0,967	0,968	0 969	0,969	0,970	0,9
1,9	0,971	0,972	0,973	0,973	0,974	0,974	0,975	0,976	0,976	0,9

Carlos Montes – uc3m

$$Z\alpha_{/2} = 1.96$$

$$\mu \in (160 \pm 1.96 \cdot {}^{10}/_{12}) = (158.36; 161.63)$$

$$\mu \in \left(160 \pm Z\alpha_{/2} \cdot \frac{10}{12}\right)$$
b)
$$1 - \alpha = 0.90 \qquad \alpha = 0.10$$

$$P\left(Z > Z\alpha_{/2}\right) = \frac{0.10}{2} = 0.05$$

$$P\left(Z < Z\alpha_{/2}\right) = 1 - 0.05 = 0.95$$

P(Z <z)< th=""><th>0,00</th><th>0,01</th><th>0,02</th><th>0,03</th><th>0,04</th><th>0,05</th><th>0,06</th><th>0,07</th><th>0,08</th><th>0,09</th></z)<>	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,500	0,504	0,508	0,512	0516	0,520	0,524	0,528	0,532	0,536
0,1	0,540	0,544	0,548	0,552	0,556	0,560	0,564	0,567	0,571	0,575
0,2	0,579	0,583	0,587	0,591	0,595	0,599	0,603	0,606	0,610	0,614
0,3	0,618	0,622	0,626	0,629	0,633	0,637	0,641	0,644	0,648	0,652
0,4	0,655	0,659	0,663	0,666	0.670	0,674	0,677	0,681	0,684	0,688
0,5	0,691	0,695	0,698	0,702	0.705	0,709	0,712	0,716	0,719	0,722
0,6	0,726	0,729	0,732	0,736	0,739	0,742	0,745	0,749	0,752	0,755
0,7	0,758	0,761	0,764	0,767	0.770	0,773	0,776	0,779	0,782	0,785
0,8	0,788	0,791	0,794	0,797	0.800	0,802	0,805	0,808	0,811	0,813
0,9	0,816	0,819	0,821	0,824	0.826	0,829	0,831	0,834	0,836	0,839
1	0,841	0,844	0,846	0,848	0.851	0,853	0,855	0,858	0,860	0,862
1,1	0,864	0,867	0,869	0,871	0.873	0,875	0,877	0,879	0,881	0,883
1,2	0,885	0,887	0,889	0,891	0.893	0,894	0,896	0,898	0,900	0,901
1,3	0,903	0,905	0,907	0,908	0 910	0,911	0,913	0,915	0,916	0,918
1,4	0,919	0,921	0,922	0,924	0 925	0,926	0,928	0,929	0,931	0,932
1,5	0,933	0,934	0,936	0,937	0 938	0,939	0,941	0,942	0,943	0,944
1,6	0,945	0,946	0,947	0,948	0,949	0,951	0,952	0,953	0,954	0,954
1,7	0,955	0,956	0,957	0,958	0,959	0,960	0,961	0,962	0,962	0,963
1,8	0,964	0,965	0,966	0,966	0,967	0,968	0,969	0,969	0,970	0,971
1,9	0,971	0,972	0,973	0,973	0,974	0,974	0,975	0,976	0,976	0,977

$$Z\alpha_{/2} = 1.64$$

$$\mu \in (160 \pm 1.64 \cdot {}^{10}/_{12}) = (158.63; 161.36)$$
(158.36; 161.63)

Carlos Montes - uc3m

2.2. Intervalo de confianza para la media con σ desconocida

$$sesgo(s^2) = E(s^2) - \sigma^2$$

$$sesgo(s^2) = \sigma^2 \frac{n-1}{n} - \sigma^2 = -\frac{\sigma^2}{n}$$

El sesgo es negativo \Rightarrow s² subestima la verdadera varianza.

2.2. Intervalo de confianza para la media con σ desconocida

 σ^2 suele ser desconocido, con lo que lo sustituimos por una estimación.

$$\hat{\sigma}^2 = s^2$$

Pero es un estimador sesgado, pudiéndose demostrar que:

$$E(s^2) = \sigma^2 \frac{n-1}{n}$$

2.2. Intervalo de confianza para la media con σ desconocida

Para corregir el sesgo:

$$E\left(s^2 \cdot \frac{n}{n-1}\right) = \frac{n}{n-1} \cdot E(s^2) = \frac{n}{n-1} \cdot \sigma^2 \cdot \frac{n-1}{n} = \sigma^2$$

Pero ¿qué es
$$\left(s^2 \cdot \frac{n}{n-1}\right)$$
?

$$\left(s^2 \cdot \frac{n}{n-1}\right) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \frac{n}{n-1} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \hat{s}^2$$

2.2. Intervalo de confianza para la media con σ desconocida

Así, si n es grande, el intervalo de confianza para la media poblacional μ es:

$$\overline{x} \pm Z_{\alpha/2} \frac{\hat{s}}{\sqrt{n}}$$

 \hat{s}^2

cuasivarianza varianza corregida

Carlos Montes - uc3m

2.2. Intervalo de confianza para la media con σ desconocida

Si la muestra no es grande, pero la población es normal:

$$\mu \in \left(\overline{x} \pm t_{n-1,\alpha/2} \cdot \hat{s} / \sqrt{n}\right)$$

2.2. Intervalo de confianza para la media con σ desconocida

Statgraphics calcula siempre estos intervalos :

- Para muestras grandes valen para cualquier población.
- Para muestras pequeñas, solo para poblaciones normales.

3. T de Student

William Gosset "Student" (1876-1937)

Definida en 1908, es la distribución:

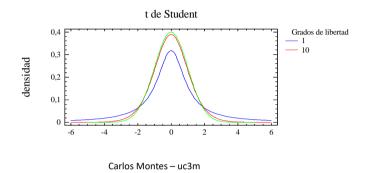


$$t_n = \frac{Z}{\sqrt{\frac{1}{n}\chi_n^2}}$$

Con $Z \rightarrow N(0,1)$

3. T de Student

Es una variable simétrica, con mayor dispersión que la normal estándar, a la que tiende rápidamente al aumentar n.



4.1. Intervalo de confianza para la varianza de poblaciones normales

$$\frac{\hat{s}^2}{\sigma^2} \rightarrow \frac{\chi_{n-1}^2}{n-1}$$
 Lema de Fisher-Cochran





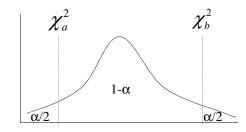
$$\frac{(n-1) \hat{s}^2}{\sigma^2} \rightarrow \chi_{n-1}^2$$

(1890-1962) (1909-1980)

4.1. Intervalo de confianza para la varianza de poblaciones normales

Llamando χ_a^2 y χ_b^2 a los valores χ_{n-1}^2 que dejan entre sí el 1- α de la distribución:

$$P\left(\chi_a^2 \le \frac{(n-1)s^2}{\sigma^2} \le \chi_b^2\right) = 1 - \alpha$$



4.1. Intervalo de confianza para la varianza de poblaciones normales

El intervalo de confianza buscado queda:

$$\frac{(n-1)\hat{s}^2}{\chi_a^2} \ge \sigma^2 \ge \frac{(n-1)\hat{s}^2}{\chi_b^2}$$

Para simplificar, se suele tomar un intervalo simétrico:

$$P(\chi_{n-1}^2 \ge \chi_b^2) = \frac{\alpha}{2}$$

$$P(\chi_{n-1}^2 \le \chi_a^2) = \frac{\alpha}{2}$$

4.1. Intervalo de confianza para la varianza de poblaciones normales

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

Carlos Montes - uc3m

ej. 48

Para estimar el porcentaje de tornillos defectuosos producidos por una máquina, se extrae una muestra de 400 unidades, de las cuales 8 resultan ser defectuosas. Construya un intervalo de confianza al 95% para el porcentaje de tornillos defectuosos fabricados.

4.2. Intervalo de confianza para una proporción

Se desea estimar la proporción p de elementos de la población que poseen un atributo determinado.

$$\overline{x} = p$$
 $\overline{x} \pm Z_{\alpha/2} \frac{\hat{s}}{\sqrt{n}}$ $\hat{s}^2 = \hat{p}\hat{q}$

$$p \in \left(\widehat{p} \pm Z\alpha_{/2}\sqrt{\frac{\widehat{p}\widehat{q}}{n}}\right)$$

$$n = 400$$
 $\hat{p} = \frac{8}{400} = 0.02$ $\hat{q} = 1 - 0.02 = 0.98$

El intervalo de confianza para el porcentaje de tornillos defectuosos es:

$$p \in \left(\hat{p} \pm Z\alpha_{/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) \quad p \in \left(0.02 \pm 1.96\sqrt{\frac{0.02 \cdot 0.98}{400}}\right)$$

$$Z\alpha_{/2}$$
=1.96 (6.28 · 10⁻³, 0.03372)

4.3. Intervalo de confianza para la λ de Poisson

Sean X_1 , X_2 ,... X_n m.a.s de una distribución de Poisson.

$$\frac{\overline{x} - \lambda}{\sqrt{s/\sqrt{n}}} \to N(0,1)$$

Como
$$\lambda = \text{var}(X) = E(X)$$

$$\frac{\overline{x} - \lambda}{\sqrt{\frac{\overline{x}}{\sqrt{n}}}} \rightarrow N(0,1)$$

Carlos Montes - uc3m

4.3. Intervalo de confianza para la λ de Poisson

Un intervalo de confianza (1- α) será aquel para el cual:

$$P\left(-Z_{\alpha/2} \le \frac{\overline{x} - \lambda}{\sqrt{\overline{x}/\sqrt{n}}} \le Z_{\alpha/2}\right) = 1 - \alpha$$

El intervalo de confianza queda:

$$\lambda \in \left(\overline{x} \pm Z \alpha_{/2} \sqrt{\frac{\overline{x}}{n}}\right)$$

5. Determinación del tamaño muestral

Función de la precisión que se guiera conseguir.

a) Tamaño muestral para la estimación de la media

$$\overline{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$
 $\overline{x} \pm L$

$$\overline{x} \pm L$$

$$L = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \Rightarrow n = \frac{Z_{\alpha/2}^2 \sigma^2}{L^2}$$

2.1. Intervalo de confianza para la media con σ conocida

b) Tamaño muestral para la estimación de una proporción

$$\hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$
 $\hat{p} \pm L$ $L = Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$

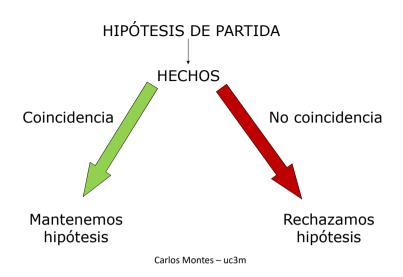
$$\hat{p}\pm L$$

$$L = Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

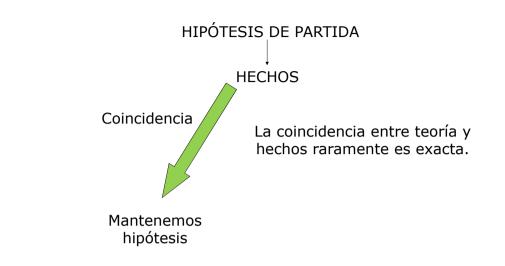
Como p es desconocido, tomamos el caso más desfavorable: p=0.5; q=0.5

$$n = \frac{Z_{\alpha/2}^2}{4L^2}$$

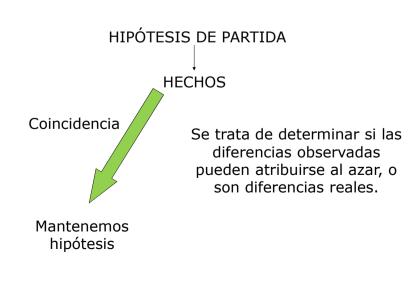
6.1. Contraste de hipótesis. Concepto



6.1. Contraste de hipótesis. Concepto



6.1. Contraste de hipótesis. Concepto



6.1. Contraste de hipótesis. Concepto

En un contraste de hipótesis se comparan 2 alternativas:

H₀: hipótesis nula o "neutra"

- La que se contrasta (la más simple).
- La que mantendremos a no ser que los datos indiquen su falsedad.
- Aceptarla no equivale a probarla, sino a no rebatirla.

6.1. Contraste de hipótesis. Concepto

H₁: hipótesis alternativa

• La que aceptamos si los datos parecen incompatibles con la hipótesis nula.

Carlos Montes – uc3m

6.1. Contraste de hipótesis. Concepto

La hipótesis nula debe tener siempre el signo =

$$H_0: \theta = \theta_0$$

Contraste bilateral

$$H_1: \theta \neq \theta_0$$

La hipótesis alternativa es lo que nos preguntamos.

$$H_0: \theta = \theta_0$$

$$H_1: \theta > \theta_0$$

o bien:
$$H_1:\theta<\theta_0$$

Contrastes unilaterales

6.2. Contraste de hipótesis. Método

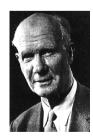
Fisher, Neyman y Pearson (1920-33)



(1890-1962)



(1894-1981)



(1895-1980)

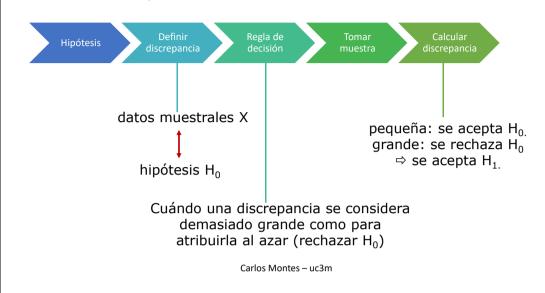
6.2. Contraste de hipótesis. Método

Su lógica es similar a la de un juicio penal:

el acusado es inocente si no se demuestra lo contrario.

la hipótesis nula es verdadera si no se demuestra lo contrario.

6.2. Contraste de hipótesis. Método

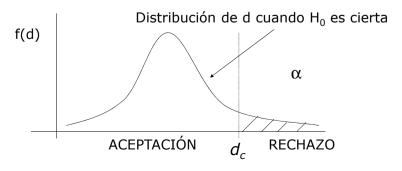


Hipótesis Definir discrepancia Regla de decisión discrepancia discrep

6.2. Contraste de hipótesis. Método

Si la discrepancia observada en la muestra, d cae dentro de la región de rechazo, rechazaremos H_0 .

En caso contrario, la aceptaremos.



6.2. Contraste de hipótesis. Método

Podemos cometer dos tipos de errores

	H _o cierta	H _o falsa
Rechazamos H _o	Error tipo I $(lpha)$	
Aceptamos H _o		Error tipo II

$$\alpha = P \left(\begin{array}{c} rechazar \ H_0 / \\ H_0 \end{array} \right) cierta$$

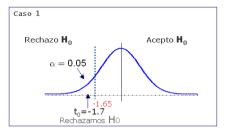
6.2. Contraste de hipótesis. Método

Objeciones:

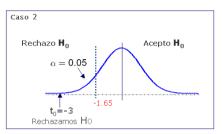
- \bullet El resultado puede ser distinto al emplear α ligeramente diferentes.
- Al emplear un α prefijado no podemos saber el grado de evidencia que la muestra indica a favor o en contra de H_0 .

Carlos Montes - uc3m

6.2. Contraste de hipótesis. Método



... en el caso 2 tenemos mayor seguridad. Aunque rechazamos H₀ en los dos casos...



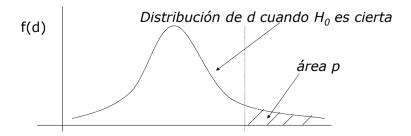
6.3. p-valor

Para determinar la región de rechazo, empleamos en lugar del nivel de significación α el nivel crítico p (p-value)

Es la probabilidad de obtener una discrepancia mayor o igual que la observada en la muestra cuando H₀ es cierta.

6.3. p-valor

• Es el mínimo nivel de significación que nos llevaría a rechazar la hipótesis nula.



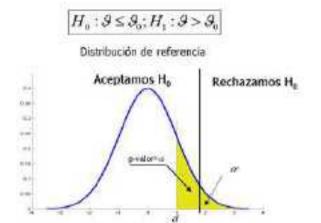
6.3. p-valor

• p<0,05 (0,01): existe muy poca evidencia en la muestra a favor de la hipótesis, luego RECHAZAMOS.

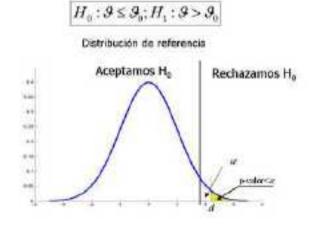
Probabilidad pequeña ⇒ discrepancia grande (supuesta H₀ cierta)

Carlos Montes - uc3m

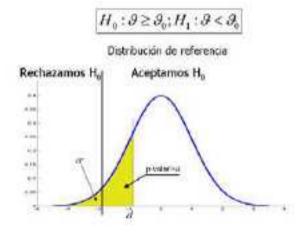
6.3. p-valor



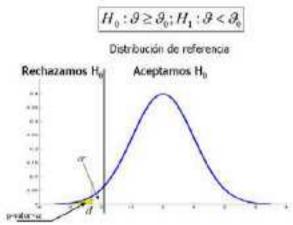
6.3. p-valor



6.3. p-valor



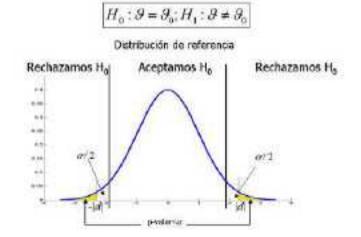
6.3. p-valor



Carlos Montes – uc3m

6.3. p-valor

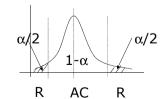
6.3. p-valor



7. Contraste para la media

Se desea contrastar la hipótesis de que la media de una distribución normal es $\mu_{\rm 0}$

$$H_{0}$$
: $\mu = \mu_{0}$ H_{1} : $\mu \neq \mu_{0}$



Contraste bilateral

7. Contraste para la media

Para cualquier variable X de media μ y varianza σ^2 , **si el** tamaño muestral es suficientemente grande se cumple que:

$$Z = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}} \to N(0, 1)$$

Y además:

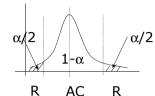
$$T = \frac{\overline{x} - \mu}{\hat{s} / \sqrt{n}} \to N(0, 1)$$

Carlos Montes – uc3m

7. Contraste para la media

Rechazamos H₀ si:

$$\frac{\overline{x} - \mu_0}{\hat{s}} \begin{cases} > Z_{\alpha/2} & \alpha/2 \\ < -Z_{\alpha/2} & R & AC & R \end{cases}$$



7. Contraste para la media

 H_0 : $\mu \leq \mu_0$ Contraste unilateral $H_1: \mu > \mu_0$

Rechazamos H_0 si: $\frac{\overline{x} - \mu_0}{\hat{s}/\sqrt{n}} > Z_{\alpha}$

7. Contraste para la media

 H_0 : $\mu \geq \mu_0$ Contraste unilateral H_1 : $\mu < \mu_0$

Rechazamos H_0 si: $\frac{\overline{x} - \mu_0}{\hat{s}/\sqrt{z}} < -Z_{\alpha}$

7. Contraste para la media

Un intervalo de confianza con nivel 1-α y un contraste usan la misma información.

La realización de un contraste de hipótesis bilateral con nivel de significación α es equivalente a realizar un intervalo de confianza de nivel $(1-\alpha)$.

Carlos Montes – uc3m