INFERENCIA ESTADÍSTICA PARA DOS POBLACIONES.

Escuela Superior
de Ingenieria

INTRODUCCIÓN

La inferencia estadística para dos poblaciones se utiliza cuando nos hacemos preguntas que tienen relación con la comparación entre los parámetros principales (medias, varianzas, proporciones…) de dos poblaciones o grupos de una población.

Se parte de la extracción de una muestra de cada población, con la que se puede construir un estadístico cuya distribución (exacta o aproximada) será conocida suponiendo determinadas condiciones (independencia, normalidad,…).

Con este estadístico se podrá realizar el contraste de comparación de los parámetros y/o construir el intervalo de confianza que estime la posible diferencia entre los parámetros que comparamos.

INTRODUCCIÓN

A continuación veremos los contrastes e intervalos de confianza para:

- Para comparar las medias de dos poblaciones o grupos
- Para comparar las varianzas de dos poblaciones o grupos
- Para comparar las proporciones de dos poblaciones o grupos

INFERENCIA PARA DOS POBLACIONES NORMALES E INDEPENDIENTES

Generalmente, en la realidad estamos interesados en realizar comparaciones entre distintas variables aleatorias. Por ejemplo, la renta media de los andaluces frente a los gallegos, la mayor o menor desigualdad de riqueza entre dos ciudades, la altura media de los hombres y las mujeres, etc. Simplificamos el estudio de la comparación de dos poblaciones a poblaciones normales e independientes.

Población 1

$$X \rightarrow N(\mu_1, \sigma_1)$$

μ_1 : Valor medio Población 1

σ_1 : Desviación Población 1

Población 2

$$X \rightarrow N(\mu_2, \sigma_2)$$

μ_2 : Valor medio Población 2

σ_2 : Desviación Población 2

INFERENCIA PARA DOS POBLACIONES NORMALES E INDEPENDIENTES

Problemas presentados:

1. Comparar μ_1 y μ_2
2. Comparar σ_1 y σ_2

Intuitivamente $\longrightarrow$ Optamos por dos estudios marginales:

1. Construir estimaciones puntuales de cada parámetro, en cada población
2. Construir un intervalo de confianza para cada parámetro, en cada población

Escuela Superior de Ingeniería

INFERENCIA PARA DOS POBLACIONES NORMALES E INDEPENDIENTES

1. Calculamos <u>estimaciones puntuales</u>:

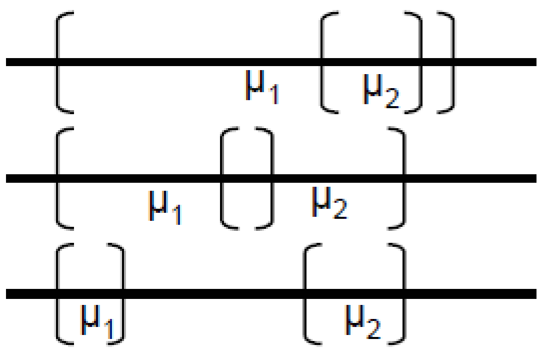
Ingresos 1 $\rightarrow N(\mu_1, \sigma_1) \rightarrow$ Estimación 1000 euros

Ingresos 2 $\rightarrow N(\mu_2, \sigma_2) \rightarrow$ Estimación 1100 euros

¿Cuál es mayor? ¿Puedo precisar?

2. Calculamos <u>intervalos de confianza</u>:

Solapamiento de intervalos:



Un intervalo $\rightarrow$ Un error

Dos intervalos

$\rightarrow$ Dos errores marginales

$\rightarrow$ Mayor error global

Presenta dos problemas

INFERENCIA PARA DOS POBLACIONES NORMALES E INDEPENDIENTES

Conclusión $\rightarrow$ Método poco preciso

Solución $\rightarrow$ Transformamos el problema en otro que permita hacer comparaciones y que sea preciso

Transformamos el problema en el estudio de un único parámetro

Comparar medias:
Parámetro resta $\theta = \mu_1 - \mu_2$

Tomamos la resta intencionadamente porque la resta de Normales es también **Normal**.

Comparar desviaciones:
Parámetro cociente $\theta = \frac{\sigma_1^2}{\sigma_2^2}$

Tomamos el cociente intencionadamente porque el cociente de Chi-cuadrados es una distribución conocida: **F de Snedecor**

INFERENCIA PARA DOS POBLACIONES NORMALES E INDEPENDIENTES

Comparación de medias:

$$\theta = 0 \rightarrow \mu_1 = \mu_2$$

$$\theta > 0 \rightarrow \mu_1 > \mu_2$$

$$\theta < 0 \rightarrow \mu_1 < \mu_2$$

Comparación de desviaciones:

$$\theta = 1 \rightarrow \sigma_1^2 = \sigma_2^2$$

$$\theta > 1 \rightarrow \sigma_1^2 > \sigma_2^2$$

$$\theta < 1 \rightarrow \sigma_1^2 < \sigma_2^2$$

INFERENCIA PARA DOS POBLACIONES NORMALES E INDEPENDIENTES

Población 1

$X_1 \to N(\mu_1, \sigma_1)$

$x_1, \ldots, x_{n_1}$ m.a.s

Tamaño muestral n_1

$\bar{x}_1$ estimación puntual de μ_1

$$\bar{x}_1 \to N\left(\mu_1, \frac{\sigma_1}{\sqrt{n_1}}\right)$$

S_1^2 estimación sesgada de σ_1^2

S_{C1}^2 estimación insegada de σ_1^2

$$\frac{n_1 S_1^2}{\sigma_1^2} = \frac{(n_1 - 1) S_{C1}^2}{\sigma_1^2} \to \chi^2_{n_1 - 1}$$

Población 2

$X_2 \to N(\mu_2, \sigma_2)$

$x_1, \ldots, x_{n_2}$ m.a.s

Tamaño muestral n_2

$\bar{x}_2$ estimación puntual de μ_2

$$\bar{x}_2 \to N\left(\mu_2, \frac{\sigma_2}{\sqrt{n_2}}\right)$$

S_2^2 estimación sesgada de σ_2^2

S_{C2}^2 estimación insegada de σ_2^2

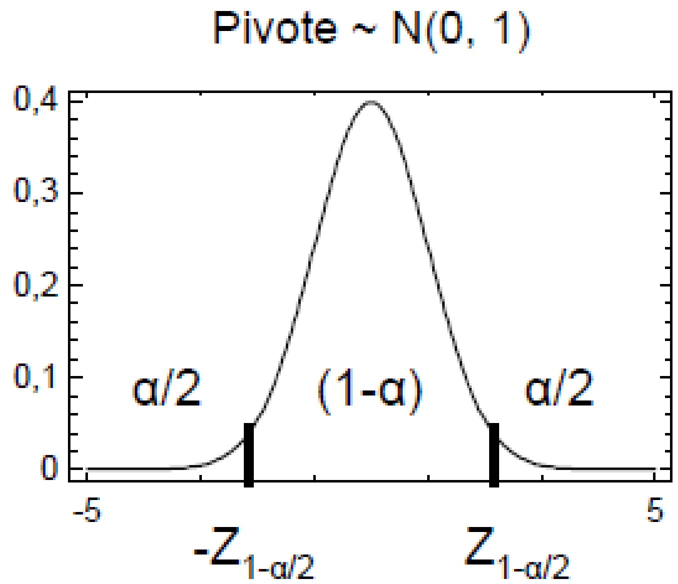
$$\frac{n_2 S_2^2}{\sigma_2^2} = \frac{(n_2 - 1) S_{C2}^2}{\sigma_2^2} \to \chi^2_{n_2 - 1}$$

Escuela Superior de Ingeniería

1. Con desviaciones típicas conocidas

$\bar{x}_1 - \bar{x}_2$ estimador de $\mu_1 - \mu_2$

$$\bar{x}_1 - \bar{x}_2 \rightarrow N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) \Rightarrow \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightarrow N(0,1)$$



INFERENCIA PARA LA DIFERENCIA DE MEDIAS

1.a. Intervalo de confianza

$$IC_{1-\alpha}(\mu_1-\mu_2)=\left((\bar{x}_1-\bar{x}_2)\pm z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

1.b. Contraste de hipótesis

$$H_0:\mu_1-\mu_2=d_0$$

Si $H_0:\mu_1-\mu_2=d_0 \Rightarrow$

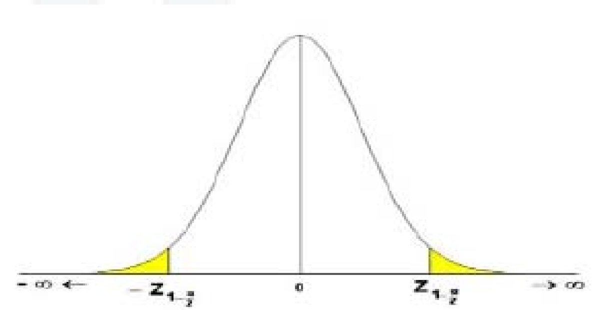
$$z_{\exp}=\frac{\bar{x}_1-\bar{x}_2-d_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\to N(0,1)$$

Estadístico de Contraste

INFERENCIA PARA LA DIFERENCIA DE MEDIAS

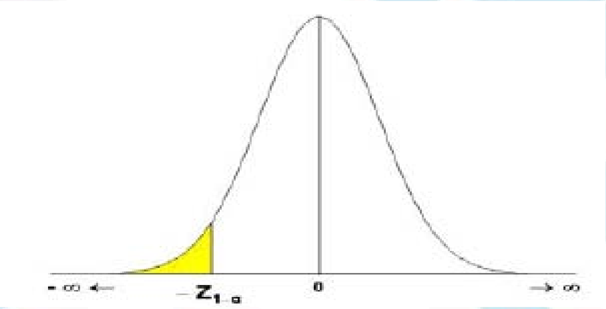
Contrastes Bilaterales:

$$H_1 : \mu_1 - \mu_2 \neq d_0 \Rightarrow \text{R.C} = \left\{ | z_{\exp} | > z_{1-\alpha/2} \right\}$$



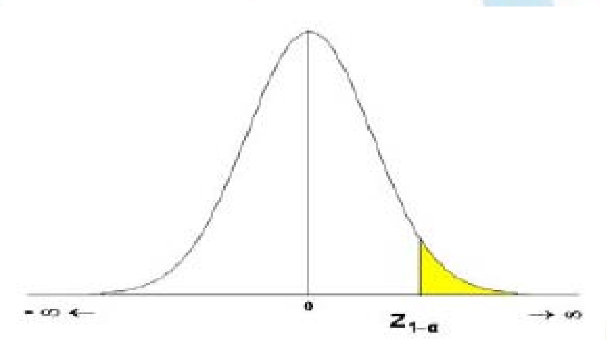
Contrastes Unilaterales:

$$H_1 : \mu_1 - \mu_2 < d_0 \Rightarrow \text{R.C} = \left\{ z_{\exp} < -z_{1-\alpha} \right\}$$



ó

$$H_1 : \mu_1 - \mu_2 > d_0 \Rightarrow \text{R.C} = \left\{ z_{\exp} > z_{1-\alpha} \right\}$$



INFERENCIA PARA LA DIFERENCIA DE MEDIAS

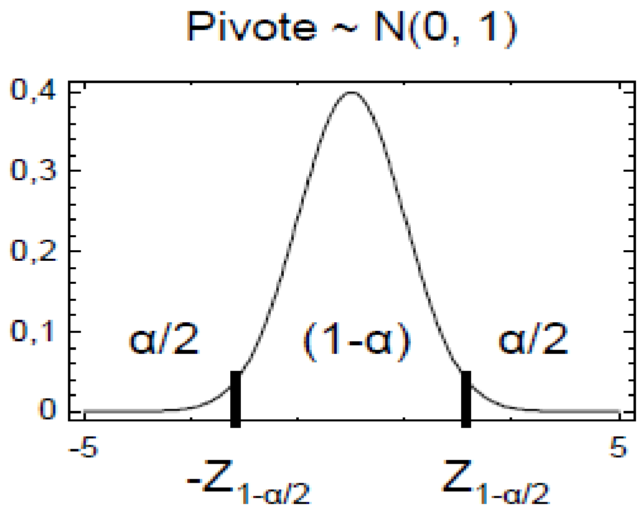
Ejemplo

Sean X e Y dos variables aleatorias normales que representan la demanda media anual de un bien de consumo en dos comunidades autónomas. Sabemos que las desviaciones típicas en ambas poblaciones son 4 y 7, respectivamente. Estamos interesados en comprobar si el consumo medio de dicho producto es idéntico en ambas comunidades con un nivel de significación del 5%. Para ello tomamos una muestra de tamaño 100 en cada una de las dos comunidades, obteniéndose medias muestrales de 40 y 45, respectivamente.

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases}$$

$$z_{\exp} = \frac{\bar{x}_1 - \bar{x}_2 - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{40 - 45 - 0}{\sqrt{\frac{4^2}{100} + \frac{7^2}{100}}} = -6{,}2017$$

Ejemplo



$$z_{1-\alpha/2} = z_{0.975} = 1.96$$

$$-z_{1-\alpha/2} = z_{\alpha/2} = z_{0.025} = -1.96$$

Rechazo H_0 si $z_{exp} \geq z_{1-\alpha/2}$ ó $z_{exp} \leq z_{\alpha/2}$

Como -6.2017 < -1.96, el valor del estadístico de contraste pertenece a la región crítica y en consecuencia **Rechazo H_0.**

Conclusión: Hay evidencias estadísticas suficientes para decir que el consumo medio es distinto en ambas comunidades con una significación del 5%.

Ejercicio propuesto.
Resolver el mismo problema mediante un intervalo de confianza.

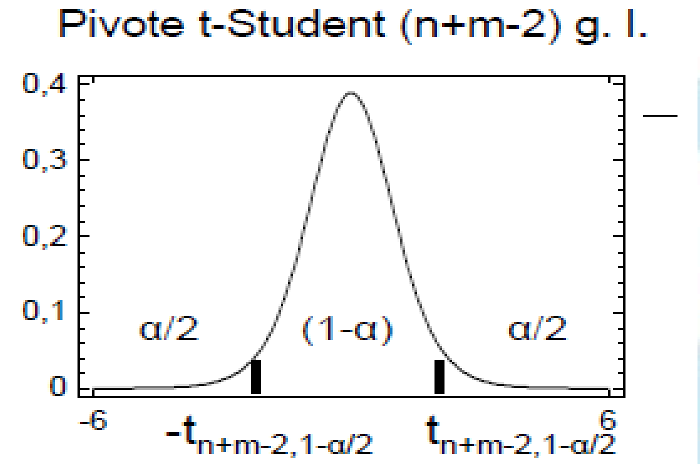
2. Con desviaciones típicas desconocidas

2.1. Con desviaciones típicas desconocidas, pero iguales

$\overline{x}_1 - \overline{x}_2$ estimador de $\mu_1 - \mu_2$

$$\overline{x}_1 - \overline{x}_2 \to N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

$\sigma_1^2 = \sigma_2^2 = \sigma^2$ desconocido



$$\Rightarrow \quad \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\sqrt{\frac{(n_1 - 1)Sc_1^2 + (n_2 - 1)Sc_2^2}{n_1 + n_2 - 2}}} \to t_{n_1 + n_2 - 2}$$

$$\boxed{\hat{S}_{CONJ}}$$

2.1.a. Intervalo de confianza

$$IC_{1-\alpha}(\mu_1-\mu_2)=\left((\overline{x}_1-\overline{x}_2)\pm t_{n_1+n_2-2;1-\alpha/2}\,\hat{S}_{CONJ}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

2.1.b. Contraste de hipótesis

$$H_0:\mu_1-\mu_2=d_0$$

Si $H_0:\mu_1 - \mu_2=d_0 \Rightarrow$

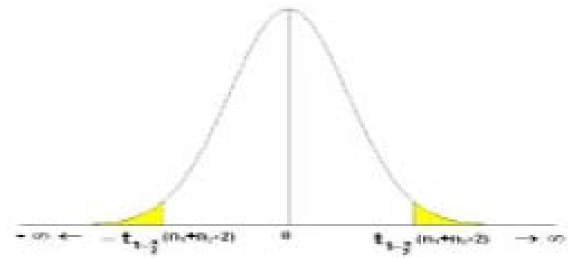
$$t_{\exp}=\frac{\overline{x}_1-\overline{x}_2-d_0}{\hat{S}_{CONJ}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\rightarrow t_{n_1+n_2-2}$$

Estadístico de Contraste

INFERENCIA PARA LA DIFERENCIA DE MEDIAS

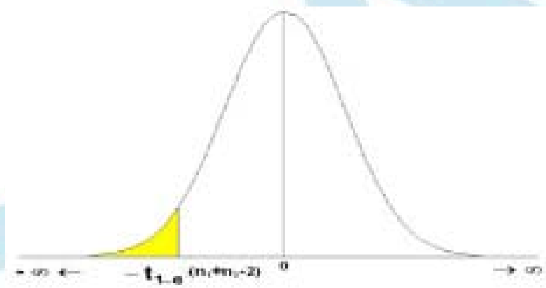
Contrastes Bilaterales:

$$\boxed{H_1 : \mu_1 - \mu_2 \neq d_0 \Rightarrow \text{ R.C} = \left\{ | t_{\exp} | > t_{n_1+n_2-2;1-\alpha/2} \right\}}$$



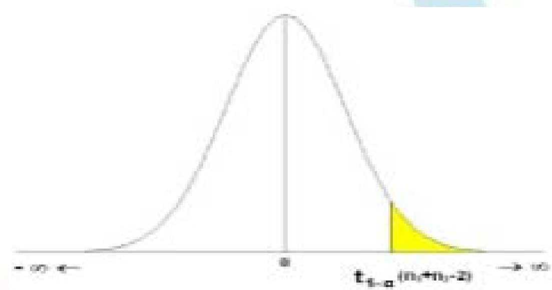
Contrastes Unilaterales:

$$\boxed{H_1 : \mu_1 - \mu_2 < d_0 \Rightarrow \text{ R.C} = \left\{ t_{\exp} < -t_{n_1+n_2-2;1-\alpha} \right\}}$$



ó

$$\boxed{H_1 : \mu_1 - \mu_2 > d_0 \Rightarrow \text{ R.C} = \left\{ t_{\exp} > t_{n_1+n_2-2;1-\alpha} \right\}}$$



INFERENCIA PARA LA DIFERENCIA DE MEDIAS

Ejemplo

Dos universidades siguen métodos distintos a la hora de matricular a sus alumnos. Se desea comparar el tiempo medio que tardan los alumnos en completar los trámites de matrícula. Cada universidad anotó dichos tiempos medios en 100 alumnos seleccionados al azar, obteniéndose:

$$\overline{x} = 50{,}2,\ Sc_x = 4.8,\ \overline{y} = 52.9,\ Sc_y = 5.4$$

Si se supone que el muestreo se llevó a cabo entre dos poblaciones distribuidas normalmente, con la misma varianza e independientes, obtener un intervalo de confianza al 90% para la diferencia entre las medias de tiempos de inscripción. ¿Podríamos pensar que existe una diferencia real entre los tiempos de ambas universidades?

Ejemplo

Como el nivel de confianza que nos piden es 90% $\longrightarrow$ $1-\alpha = 0.9$ $\longrightarrow$ $\alpha/2 = 0.05$ $\longrightarrow$ $1 - \alpha/2 = 0.95$

Por otro lado $n_1 + n_2 - 2 = 100 + 100 - 2 = 198$

Con lo que, el percentil de la t de Student que tenemos que utilizar es $t_{198;0.95} = 1.65$.

$$\hat{S}_{CONJ} = \sqrt{\frac{99 \cdot 4.8^2 + 99 \cdot 5.4^2}{100 + 100 - 2}} = 5.11$$

Por tanto, el intervalo es:

$$IC_{90\%}(\mu_1 - \mu_2) = \left((50.2 - 52.9) \pm 1.65 \cdot 5.11 \sqrt{1/100 + 1/100} \right) = (-3.89, -1.50)$$

Solución: Como el intervalo no incluye el 0, existen evidencias estadísticas suficientes para pensar que los tiempos serán diferentes con un 90% de confianza.

Ejercicio propuesto.
Resolver el mismo problema mediante un contraste de hipótesis.

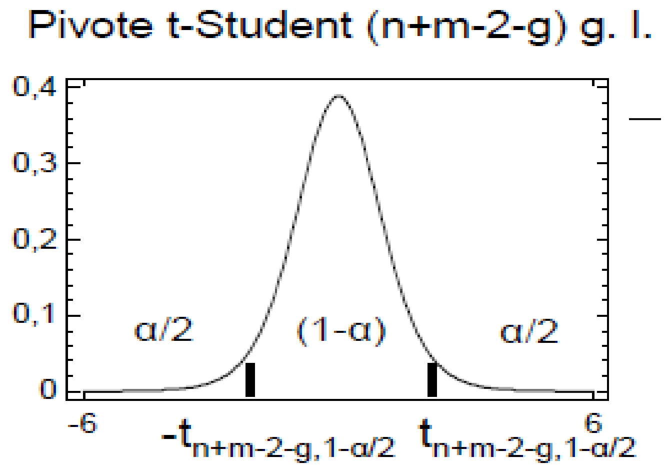
2. Con desviaciones típicas desconocidas

2.1. Con desviaciones típicas desconocidas y distintas

$\overline{x}_1 - \overline{x}_2$ estimador de $\mu_1 - \mu_2$

$$\overline{x}_1 - \overline{x}_2 \rightarrow N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

$\sigma_1^2 \neq \sigma_2^2$ y desconocidos



$$\Rightarrow \quad \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{Sc_1^2}{n_1} + \frac{Sc_2^2}{n_2}}} \rightarrow t_g$$

$$\text{con } \; g = \text{Redondear}\ \frac{\left(\frac{Sc_1^2}{n_1} + \frac{Sc_2^2}{n_2}\right)^2}{\frac{\frac{Sc_1^1}{n_1}}{n_1 - 1} + \frac{\frac{Sc_2^2}{n_2}}{n_2 - 1}}$$

2.2.a. Intervalo de confianza

$$IC_{1-\alpha}(\mu_1-\mu_2)=\left((\overline{x}_1-\overline{x}_2)\pm t_{g;1-\alpha/2}\sqrt{\frac{Sc_1^2}{n_1}+\frac{Sc_2^2}{n_2}}\right)$$

2.2.b. Contraste de hipótesis

$$H_0:\mu_1-\mu_2=d_0$$

Si $H_0:\mu_1-\mu_2=d_0 \Rightarrow$

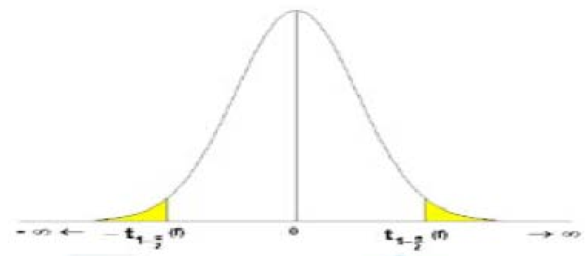
$$t_{\exp}=\frac{\overline{x}_1-\overline{x}_2-d_0}{\sqrt{\frac{Sc_1^2}{n_1}+\frac{Sc_2^2}{n_2}}}\rightarrow t_g$$

Estadístico de Contraste

INFERENCIA PARA LA DIFERENCIA DE MEDIAS

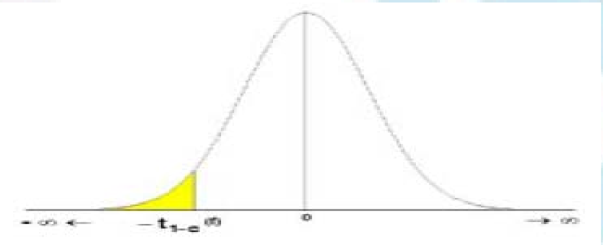
Contrastes Bilaterales:

$$H_1 : \mu \neq \mu_0 \;\Rightarrow\; \mathrm{R.C} = \left\{ |t_{\exp}| \; > t_{g;1-\alpha/2} \right\}$$



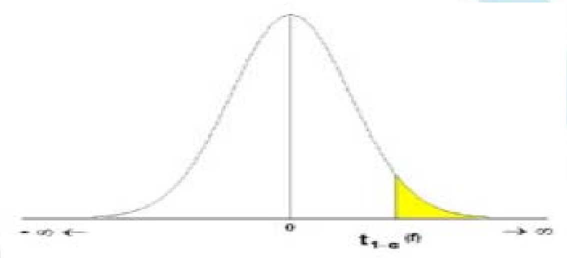
Contrastes Unilaterales:

$$H_1 : \mu < \mu_0 \;\Rightarrow\; \mathrm{R.C} = \left\{ t_{\exp} < -t_{g;1-\alpha} \right\}$$



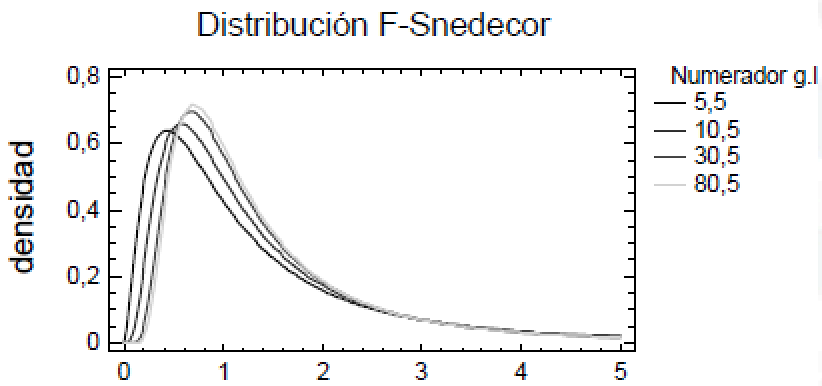
ó

$$H_1 : \mu > \mu_0 \;\Rightarrow\; \mathrm{R.C} = \left\{ t_{\exp} > t_{g;1-\alpha} \right\}$$



Escuela Superior de Ingeniería

Distribución F de Snedecor con n y m grados de libertad.



$$E[F] = {}^{m}\!/_{m-2}$$

$$Var[F] = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$$

Si tenemos dos variables independientes X e Y, que se distribuyen según una Chi-cuadrado con n y m grados de libertad, respectivamente, entonces:

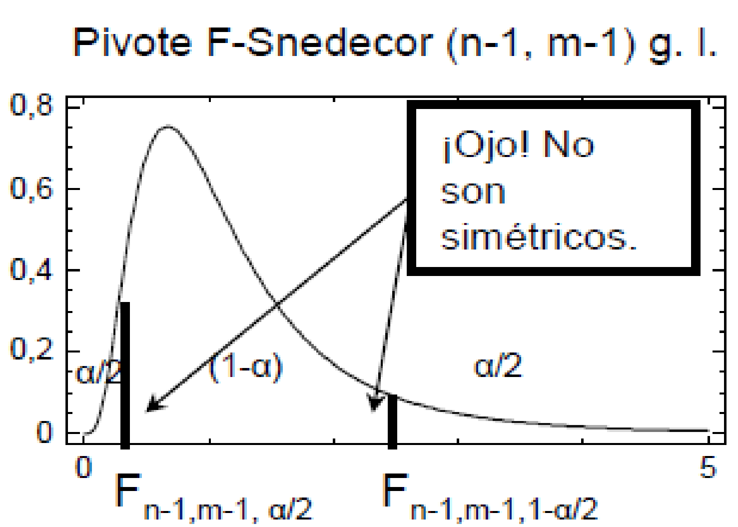
$$F_{n,m} = \frac{\chi_n^2 / n}{\chi_m^2 / m}$$

INFERENCIA PARA EL COCIENTE DE VARIANZAS

Distribución F de Snedecor con n y m grados de libertad.

Como no es una distribución simétrica, para calcular percentiles, haremos uso de la fórmula:

$$\boxed{F_{n,m;\alpha} = \frac{1}{F_{m,n;1-\alpha}}}$$



$$\boxed{\frac{\sigma_2^2}{\sigma_1^2}\frac{S_{C1}^2}{S_{C2}^2} \rightarrow F_{n_1-1,n_2-1}}$$

Escuela Superior de Ingeniería

INFERENCIA PARA EL COCIENTE DE VARIANZAS

a. Intervalo de confianza

$$IC_{1-\alpha}(\sigma_1^2/\sigma_2^2) = \left(\frac{1}{F_{n_1-1,n_2-1;1-\alpha/2}} \cdot \frac{S_{C1}^2}{S_{C2}^2}, \frac{1}{F_{n_1-1,n_2-1;\alpha/2}} \cdot \frac{S_{C1}^2}{S_{C2}^2} \right)$$

b. Contraste de hipótesis

$$H_0 : \sigma_1^2 = \sigma_2^2$$

Si $H_0 : \sigma_2^2 / \sigma_1^2 = 1 \Rightarrow$

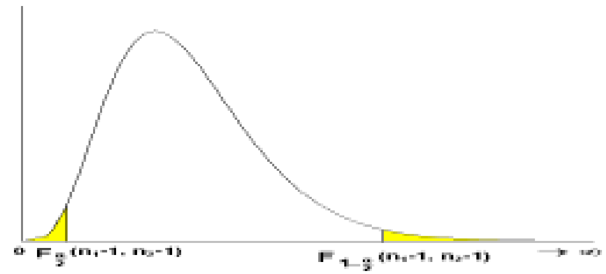
$$F_{\exp} = \frac{S_{C1}^2}{Sc_2^2} \to F_{n_1-1,n_2-1}$$

Estadístico de Contraste

INFERENCIA PARA EL COCIENTE DE VARIANZAS

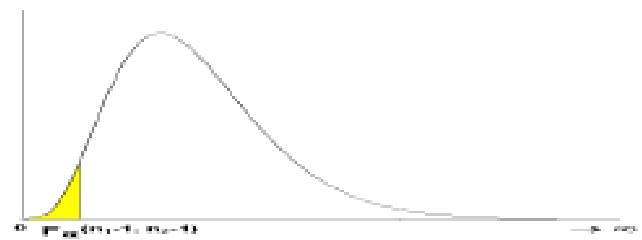
Contrastes Bilaterales:

$$H_1 : \sigma_1^2 \neq \sigma_2^2 \Rightarrow \text{ R.C} = \left\{ F_{\exp} < \text{F}_{n_1-1, n_2-1;\ \alpha/2} \right\} \cup \left\{ F_{\exp} > \text{F}_{n_1-1, n_2-1; 1-\alpha/2} \right\}$$



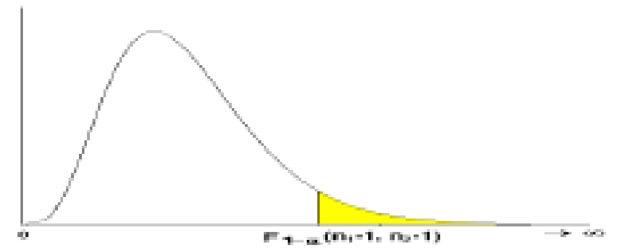
Contrastes Unilaterales:

$$H_1 : \sigma_1^2 < \sigma_2^2 \Rightarrow \text{ R.C} = \left\{ F_{\exp} < \text{F}_{n_1-1, n_2-1;\ \alpha} \right\}$$



ó

$$H_1 : \sigma_1^2 > \sigma_2^2 \Rightarrow \text{ R.C} = \left\{ F_{\exp} > \text{F}_{n_1-1, n_2-1;\ 1\text{-}\alpha} \right\}$$



INFERENCIA PARA EL COCIENTE DE VARIANZAS

Ejemplo

Para realizar un estudio histórico sobre los salarios anuales pagados por una empresa a sus empleados (en miles de euros), se selecciona al azar una m.a.s. de 5 hombres y otra de 4 mujeres, siendo el salario medio de los hombres 35,121 euros con una cuasivarianza de 5,3 y el salario medio de las mujeres 32,513 euros con una cuasivarianza 5,7. A la vista de los datos y admitiendo normalidad en la distribución de salarios, contrastar la hipótesis de igualdad de salarios. Previamente, contrastar la hipótesis de igualdad de varianzas. Utilizar una significación del 10%.

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

$$F_{\exp} = \frac{S_{C1}^2}{S_{C2}^2} = \frac{5.3}{5.7} = 0.929$$

Ejemplo

$$F_{u,v;p} = F_{4,3;0.95} = 9.12$$

$$F_{4,3;0.05} = \frac{1}{F_{3,4;0.95}} = \frac{1}{6.59} = 0.15$$

v \ u	1	2	3	4	5	6	7	8	9	10	11	12	p
1	39,863	49,500	53,593	55,833	57,240	58,204	58,906	59,439	59,858	60,195	60,473	60,705	0,900
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	242,98	243,91	0,950
1	647,79	799,50	864,16	899,58	921,85	937,11	948,22	956,66	963,28	968,63	973,03	976,71	0,975
1	4052,2	4999,5	5403,4	5624,6	5763,6	5859,0	5928,4	5981,1	6022,5	6055,8	6083,3	6106,3	0,990
1	16211	19999	21615	22500	23056	23437	23715	23925	24091	24224	24334	24426	0,995
2	8,526	9,000	9,162	9,243	9,293	9,326	9,349	9,367	9,381	9,392	9,401	9,408	0,900
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,371	19,385	19,396	19,405	19,413	0,950
2	38,506	39,000	39,165	39,248	39,298	39,331	39,355	39,373	39,387	39,398	39,407	39,415	0,975
2	98,503	99,000	99,166	99,249	99,299	99,333	99,356	99,374	99,388	99,399	99,408	99,416	0,990
2	198,50	199,00	199,17	199,25	199,30	199,33	199,36	199,37	199,39	199,40	199,41	199,42	0,995
3	5,538	5,462	5,391	5,343	5,309	5,285	5,266	5,252	5,240	5,230	5,222	5,216	0,900
3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,786	8,763	8,745	0,950
3	17,443	16,044	15,439	15,101	14,885	14,735	14,624	14,540	14,473	14,419	14,374	14,337	0,975
3	34,116	30,817	29,457	28,710	28,237	27,911	27,672	27,489	27,345	27,229	27,133	27,052	0,990
3	55,552	49,799	47,467	46,195	45,392	44,838	44,434	44,126	43,882	43,686	43,524	43,387	0,995
4	4,545	4,325	4,191	4,107	4,051	4,010	3,979	3,955	3,936	3,920	3,907	3,896	0,900
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964	5,936	5,912	0,950
4	12,218	10,649	9,979	9,605	9,364	9,197	9,074	8,980	8,905	8,844	8,794	8,751	0,975
4	21,198	18,000	16,694	15,977	15,522	15,207	14,976	14,799	14,659	14,546	14,452	14,374	0,990
4	31,333	26,284	24,259	23,155	22,456	21,975	21,622	21,352	21,139	20,967	20,824	20,705	0,995
5	4,060	3,780	3,619	3,520	3,453	3,405	3,368	3,339	3,316	3,297	3,282	3,268	0,900

Región Crítica: $F_{exp} < 0.15$ ó $F_{exp} > 9.12$

Como $0.15 < 0.929 < 9.12$ **Acepto H_0**

Conclusión: No existen evidencias estadísticas suficientes para rechazar la igualdad de varianzas poblacionales.

Ejemplo

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases}$$

$$\hat{S}_{CONJ} = \sqrt{\frac{4 \cdot 5,3 + 3 \cdot 5,7}{5 + 4 - 2}} = 2.339$$

$$t_{\exp} = \frac{\overline{x}_1 - \overline{x}_2 - d_0}{\hat{S}_{CONJ}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{35,121 - 23,513 - 0}{2,339\sqrt{\frac{1}{5} + \frac{1}{4}}} = 7.398$$

Región crítica: $t_{exp} > t_{7;0.95}$ ó $t_{exp} < t_{7;0.05}$

$t_{7;0.95} = 1.895$

$t_{7;0.05} = -1.895$

Como $7.398 > 1.895$ $\Longrightarrow$ **Rechazo H_0**

Conclusión: Hay evidencias estadísticas suficientes para afirmar que los salarios no son iguales, con una significación del 10%.

INFERENCIA PARA DOS POBLACIONES NORMALES Y DEPENDIENTES (PAREADAS). DIFERENCIA DE MEDIAS.

Tomadas las muestras $x_1,\ldots,x_n$ y $x_1^{'},\ldots,x_n^{'}$, se calculan las diferencias $d_i = x_{1i} - x_{1i}^{'}$ y se utiliza el resultado en el muestreo:

$$\bar{d} \to N\left(\mu, {}^{S_d}\!/\!_{\sqrt{n}}\right) \Rightarrow \frac{\bar{d}-\mu}{S_d/\sqrt{n}} \to t_{n-1}$$

Siendo $\bar{d}$ la media y S_d la cuasidesviación de las diferencias.

a. Intervalo de confianza

$$\boxed{IC_{1-\alpha}(\mu_1-\mu_2) = \left(\bar{d} \pm t_{n_1-1;1-\alpha/2}\frac{S_d}{\sqrt{n}}\right)}$$

INFERENCIA PARA DOS POBLACIONES NORMALES Y DEPENDIENTES (PAREADAS). DIFERENCIA DE MEDIAS.

b. Contraste de hipótesis

$$H_0 : \mu_1 - \mu_2 = d_0$$

Si $H_0 : \mu_1 - \mu_2 = d_0 \Rightarrow$

$$t_{\exp} = \frac{\bar{d} - d_0}{S_d / \sqrt{n}} \rightarrow t_{n-1}$$

Estadístico de Contraste

Las regiones críticas son las mismas que las del contraste para la media cuando la varianza es desconocida.

INFERENCIA PARA DOS POBLACIONES NORMALES Y DEPENDIENTES (PAREADAS). DIFERENCIA DE MEDIAS.

Ejemplo

Diez ratas fueron expuestas a condiciones que simulaban una enfermedad. El número de latidos por minuto antes y después del experimento fue:

Antes: 70, 84, 88, 110, 105, 100, 110, 67, 79, 86
Después: 115, 148, 176, 191, 158, 178, 179, 140, 161, 157

¿Proporcionan estos datos suficientes evidencias estadísticas para indicar que en las condiciones del experimento aumenta el número de latidos por minuto de las ratas?

Identificamos que son muestras pareadas, porque ambas son medidas de los mismos individuos en instantes distintos de tiempo.

Ejemplo

Como son muestras pareadas, calculamos las diferencias:
d_i=Antes-Después: -45, -64, -88, -81, -53, -78, -69, -73, -82, -71

Obtenemos unos parámetros muestrales para estas diferencias iguales a:

$$\bar{d} = -70.4$$

$$S_d = 13.38$$

Planteamos el siguiente contraste:

$$\begin{cases} H_0 : \mu_1 \geq \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$$ que equivalentemente es: $$\begin{cases} H_0 : \bar{d} = 0 \\ H_1 : \bar{d} < 0 \end{cases}$$

Ejemplo

Calculamos el estadístico de contraste:

$$t_{exp} = \frac{-70.4}{13.38/\sqrt{10}} = -16.64$$

Y la región crítica es $t_{exp} < t_{9;0.05} = -1.83$

Como -16.64 < -1.83, el valor del estadístico de contraste pertenece a la región crítica y en consecuencia, **Rechazo H_0.**

Conclusión: Hay evidencias estadísticas suficientes para afirmar que los latidos han aumentado después del experimento, con una significación del 5%.

Hagamos esta misma comprobación mediante un intervalo de confianza…

Ejemplo

$$IC_{95\%}(\mu_1-\mu_2)=\left(\bar{d}\pm t_{n-1;1-\alpha/2}\frac{S_d}{\sqrt{n}}\right)=\left(-70.4\pm 2.26\frac{13.38}{\sqrt{10}}\right)$$

$$IC_{95\%}(\mu_1-\mu_2)=(-70.4\pm 9.56)=(-79.96,-60.84)$$

Como el intervalo no incluye el 0, existen evidencias estadísticas suficientes para afirmar que el número de latidos por minuto no puede ser el mismo antes que después del experimento. Como además, los extremos del intervalo son ambos negativos, esto indicaría que el número de latidos después del experimento es mayor que el número de latidos que se registraban antes, con un 95% de confianza.

INFERENCIA PARA DOS POBLACIONES BINOMIALES E INDEPENDIENTES

Generalmente estamos interesados en comparar dos proporciones.

Población 1

$X_1 \rightarrow B(n_1, p_1)$

$x_1, \ldots, x_{n_1}$ m.a.s

Tamaño muestral n_1

p_1 estimación puntual de π_1

$$\mathrm{p}_1 \rightarrow N\left(\pi_1, \sqrt{\frac{\mathrm{p}_1(1-\mathrm{p}_1)}{n_1}}\right)$$

Población 2

$X_2 \rightarrow B(n_2, p_2)$

$x_1, \ldots, x_{n_2}$ m.a.s

Tamaño muestral n_2

p_2 estimación puntual de π_2

$$\mathrm{p}_2 \rightarrow N\left(\pi_2, \sqrt{\frac{\mathrm{p}_2(1-\mathrm{p}_2)}{n_2}}\right)$$

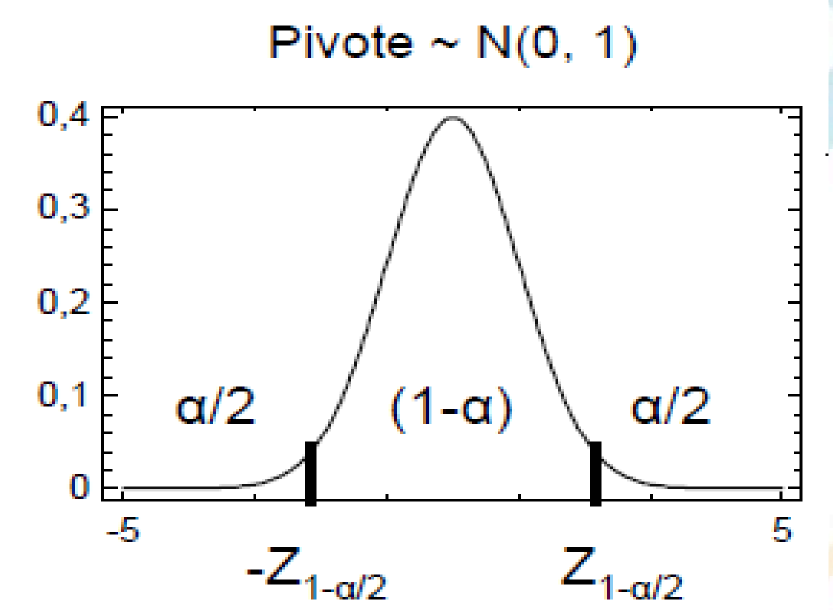
Por los mismos motivos mostrados en el caso de dos poblaciones normales, la comparación de ambas proporciones se realiza a través del parámetro resta

INFERENCIA PARA LA DIFERENCIA DE PROPORCIONES

$\mathrm{p}_1 - \mathrm{p}_2$ estimador de $\pi_1 - \pi_2$

$$\mathrm{p}_1 - \mathrm{p}_2 \rightarrow N\left(\pi_1 - \pi_2, \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}} \right)$$

$$\Rightarrow \frac{(\mathrm{p}_1 - \mathrm{p}_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{\mathrm{p}_1(1-\mathrm{p}_1)}{n_1} + \frac{\mathrm{p}_2(1-\mathrm{p}_2)}{n_2}}} \rightarrow N(0,1)$$



INFERENCIA PARA LA DIFERENCIA DE PROPORCIONES

a. Intervalo de confianza

$$IC_{1-\alpha}(\pi_1 - \pi_2) = \left((p_1 - p_2) \pm Z_{1-\alpha/2} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \right)$$

b. Contraste de hipótesis

$$H_0 : \pi_1 - \pi_2 = d_0$$

Si $H_0 : \pi_1 - \pi_2 = d_0 \Rightarrow$

$$z_{exp} = \frac{p_1 - p_2 - d_0}{\sqrt{P_{CONJ}(1-P_{CONJ})\frac{1}{n_1} + \frac{1}{n_2}}} \rightarrow N(0,1)$$

siendo

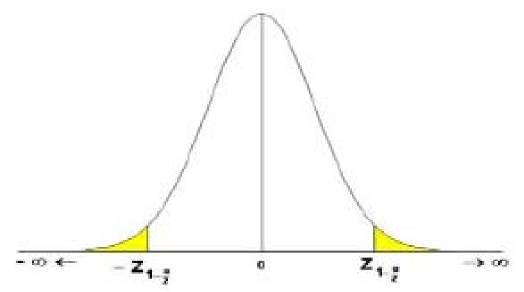
$$P_{CONJ} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

Estadístico de Contraste

INFERENCIA PARA LA DIFERENCIA DE PROPORCIONES

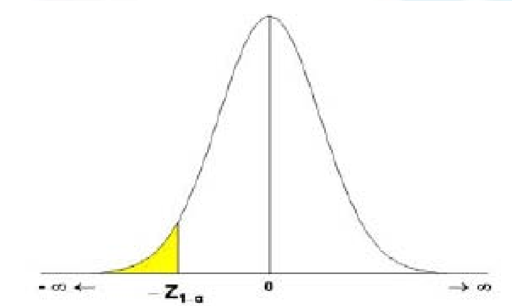
Contrastes Bilaterales:

$$H_1 : \pi \neq \pi_0 \Rightarrow \text{R.C} = \left\{ |z_{\text{exp}}| > z_{1-\alpha/2} \right\}$$



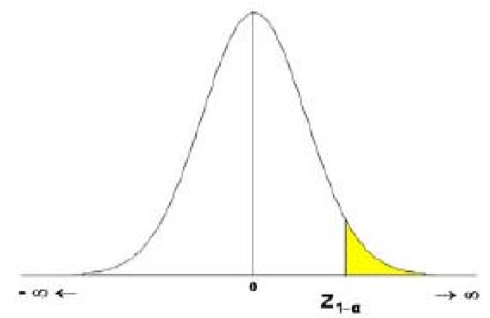
Contrastes Unilaterales:

$$H_1 : \pi < \pi_0 \Rightarrow \text{R.C} = \left\{ z_{\text{exp}} < -z_{1-\alpha} \right\}$$



ó

$$H_1 : \pi > \pi_0 \Rightarrow \text{R.C} = \left\{ z_{\text{exp}} > z_{1-\alpha} \right\}$$



Escuela Superior de Ingeniería

INFERENCIA PARA LA DIFERENCIA DE PROPORCIONES

Ejemplo

En un estudio de los hábitos de fumador para personas zurdas y diestras, se eligió una muestra aleatoria de 400 zurdos, de los cuales 190 fumaban, y una muestra aleatoria de 800 diestros, de los cuales 300 fumaban. Contraste la hipótesis de que la proporción de fumadores es la misa en diestros y zurdos para un nivel de significación α=0.01.

$$\begin{cases} H_0 : \pi_1 - \pi_2 = 0 \ (\pi_1 = \pi_2) \\ H_1 : \pi_1 - \pi_2 \neq 0 \ (\pi_1 \neq \pi_2) \end{cases}$$

$$p_1 = \frac{190}{400} = 0.475$$

$$p_2 = \frac{300}{800} = 0.375$$

Escuela Superior de Ingeniería

Ejemplo

$$P_{CONJ} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{400 \cdot 0.475 + 800 \cdot 0.375}{400 + 800} = 0.4083$$

$$z_{\exp} = \frac{0.475 - 0.375 - 0}{\sqrt{0.4083(1 - 0.4083)\left(\frac{1}{400} + \frac{1}{800}\right)}} = 3.3223$$

$$\text{Para } \alpha = 0.01 \Rightarrow z_{0.995} = 2.58 \text{ y } z_{0.005} = -2.58$$

Como 3.3223 > 2.58, el valor del estadístico de contraste pertenece a la región crítica y en consecuencia **Rechazo H_0.**

Conclusión: Hay evidencias estadísticas suficientes para decir que la proporción de fumadores no es la misma en zurdos y diestros, con una significación del 5%..

Escuela Superior de Ingeniería

EJERCICIOS PROPUESTOS

Problema 1.

Se realizó un muestreo para decidir si los sueldos de los peones de albañil de una ciudad A y de otra B son iguales en promedio o no. Para ello, se consultó a 100 peones de la ciudad A y a 150 de la ciudad B. Analizadas las respuestas dadas por dichos operarios, se determinó que la media de los sueldos de los peones de la ciudad A era de 760 euros y la de los de la ciudad B era 720 euros. Suponiendo que la desviación típica poblacional de los sueldos de A es 12 euros y la de B 9 euros, decidir si el sueldo medio en ambas ciudades es igual o distinto.

Problema 2.

Se desea comparar el gasto medio mensual en alimentación entre las familias de dos barrios. Para ello, se seleccionan 20 familias de cada barrio, observando sus gastos mensuales en alimentación. Se determinó la media y las cuasidesviaciones típicas, obteniéndose una media de 200 y 175 euros y una cuasidesviación de 20 y 17, respectivamente. Suponiendo que los gastos se distribuyen normalmente, decidir si los gastos medios en ambos barrios pueden considerarse iguales.

EJERCICIOS PROPUESTOS

Problema 3.

Se llevó a cabo un estudio para determinar el grado en el cual el alcohol entorpece la habilidad de pensamiento para llevar a cabo una determinada tarea. Se seleccionaron 10 personas al azar de distintas características para participar en el experimento. Estas personas realizaron la tarea, sin nada de alcohol en su organismo. Posteriormente, la volvieron a realizar con un contenido de alcohol en su organismo de 0.1%. Los tiempos en minutos que se tardó en realizar la tarea antes y después son:

Antes: 27, 24, 52, 45, 32, 35, 40, 30, 37, 20

Después: 39, 45, 71, 61, 46, 58, 51, 34, 48, 30

Con una significación del 5%, ¿podemos afirmar que el consumo de alcohol influye en la realización de la tarea?