9 Estimación y Decisión

ESTIMACIÓN PUNTUAL Y POR INTERVALOS

Los parámetros de la población se desconocen generalmente, siendo uno de los primordiales principios de la estadística el estimarlos. El método o estadístico empleado para ello se denomina estimador del parámetro poblacional.

Hablamos de ***estimación puntual*** cuando asignamos al parámetro de la población un valor concreto. Los mejores estimadores puntuales de la media y varianza poblacional son :

De la media poblacional : la media muestral

$$Est(\mu) = \bar{x}$$

De la varianza poblacional : la cuasivarianza muestral

$$Est(\sigma^2) = s'^2 = \frac{n}{n-1}.s^2$$

Hablamos de ***estimación por intervalos de confianza*** cuando asignamos al parámetro un intervalo de valores.

Si estimamos el parámetro **h** mediante el estimador $\bar{h}$ y, por ejemplo, tal estimador se distribuye normalmente con media $E(\bar{h}) = h$ (estimador insesgado o centrado) y desviación típica $D(\bar{h})$, sabemos que

$$Pr(-2'58 < \frac{\bar{h} - h}{D(\bar{h})} < 2'58) = 0'99 \qquad \text{o bien} \qquad Pr(\bar{h} - 2'58.D(\bar{h}) < h < \bar{h} + 2'58.D(\bar{h})) = 0'99$$

La probabilidad de que el parámetro de la población se encuentre en un cierto intervalo (0'99 en el supuesto anterior), se denomina ***nivel de confianza*** y se representa por 1-α. El valor α (0'01 o 1%) es el ***margen de error*** que se permite en la estimación.

Los intervalos de confianza de los parámetros usuales se enumeran junto a las funciones de decisión correspondientes, en el tercer epígrafe del presente tema.

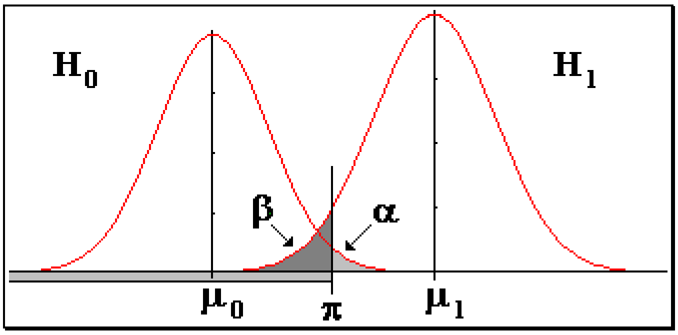
CONTRASTES PARAMÉTRICOS

Contraste es el procedimiento por el cuál decidimos si una propuesta sobre la población puede aceptarse o no.
Las notaciones y terminologías empleadas se resumen a continuación :

H_0 **Hipótesis nula**
H_1 **Hipótesis alternativa**

Error de Tipo I
Rechazar H_0 siendo cierta
α = Prob (Error de Tipo I)
= nivel de significación

Error de Tipo II
Aceptar H_0 siendo falsa
β = Prob (Error de Tipo II)



Relación entre α y β y el tamaño muestral :

$$n = \frac{\sigma^2.\left(|z_\alpha| + |z_\beta|\right)^2}{(\mu_0 - \mu_1)^2} \quad \text{unilateral}$$

$$n = \frac{\sigma^2.\left(|z_{\alpha/2}| + |z_\beta|\right)^2}{(\mu_0 - \mu_1)^2} \quad \text{bilateral}$$

Potencia del contraste	η = 1 - β
Región de aceptación	Intervalo en el que se acepta la hipótesis nula
Región de rechazo	Intervalo en el que no se acepta la hipótesis nula
Nivel(es) crítico(s) de decisión	Extremos del intervalo o región de aceptación
Función de decisión	Estadístico que permite resolver el contraste

H_0	H_1	Tipo	Región de aceptación
$h = h_0$	$h = h_1$	Unilateral	(-∞ , p) si $h_0 < h_1$ (p , +∞) si $h_0 > h_1$
$h = h_0$	$h \neq h_0$	Bilateral	(p_1 , p_2)
$h = h_0$	$h < h_0$	Unilateral	(p , +∞)
$h = h_0$	$h > h_0$	Unilateral	(-∞ , p)
$h \geq h_0$	$h < h_0$	Unilateral	(p , +∞)
$h \leq h_0$	$h > h_0$	Unilateral	(-∞ , p)

FUNCIONES DE DECISIÓN E INTERVALOS DE CONFIANZA

Enumeramos a continuación las funciones de decisión y las regiones de aceptación a emplear en los contrastes de los parámetros usuales (***decisión paramétrica***), así como los intervalos de confianza de los mismos (***estimación***).

Se sigue la notación siguiente :

	POBLACIÓN
μ	Media
σ^2	Varianza
π	Proporción

	MUESTRA
n	Tamaño de la muestra
$\bar{x}$	Media
s^2	Varianza
s'^2	Cuasivarianza
p	Proporción
r	Coeficiente de correlación

1. MEDIA

1.1. Varianza de la población conocida :

Función de decisión :	Región de aceptación (bilateral) :	Intervalo de confianza :
$z_\alpha = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$	$\mu \pm z_\alpha.\dfrac{\sigma}{\sqrt{n}}$	$\bar{x} \pm z_\alpha.\dfrac{\sigma}{\sqrt{n}}$

1.2. Varianza de la población desconocida y muestras grandes (n > 51) :

Función de decisión :	Región de aceptación (bilateral) :	Intervalo de confianza :
$z_\alpha = \dfrac{\bar{x}-\mu}{s/\sqrt{n-1}}$	$\mu \pm z_\alpha.\dfrac{s}{\sqrt{n-1}}$	$\bar{x} \pm z_\alpha.\dfrac{s}{\sqrt{n-1}}$

1.3. Varianza de la población desconocida y muestras pequeñas (n ≤ 51) :

Función de decisión :	Región de aceptación (bilateral) :	Intervalo de confianza :
$t_{n-1,\alpha} = \dfrac{\bar{x}-\mu}{s/\sqrt{n-1}}$	$\mu \pm t_{n-1,\alpha}.\dfrac{s}{\sqrt{n-1}}$	$\bar{x} \pm t_{n-1,\alpha}.\dfrac{s}{\sqrt{n-1}}$

2. VARIANZA

Función de decisión :	Región de aceptación (bilateral) :	Intervalo de confianza :
$\chi^2_{n-1,\alpha} = \dfrac{n.s^2}{\sigma^2}$	$\left(\dfrac{\sigma^2.\chi_1^2}{n}, \dfrac{\sigma^2.\chi_2^2}{n}\right)$	$\left(\dfrac{n.s^2}{\chi_1^2}, \dfrac{n.s^2}{\chi_2^2}\right)$

3. DIFERENCIAS DE MEDIAS

3.1. Muestras independientes y varianzas de las poblaciones conocidas :

Función de decisión :	Región de aceptación (bilateral) :	Intervalo de confianza :
$z_\alpha = \dfrac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$	$(\mu_1-\mu_2) \pm z_\alpha.\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$	$(\bar{x}_1-\bar{x}_2) \pm z_\alpha.\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$

3.2. Muestras independientes y varianzas de las poblaciones desconocidas pero iguales :

A) Muestras grandes (n_1+n_2-2 > 50) :

Función de decisión :	Región de aceptación (bilateral) :
$z_\alpha = \dfrac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{n_1.s_1^2+n_2.s_2^2}{n_1+n_2-2}}.\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$	$(\mu_1-\mu_2) \pm z_\alpha.\sqrt{\dfrac{n_1.s_1^2+n_2.s_2^2}{n_1+n_2-2}}.\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}$

	Intervalo de confianza : $\left(\overline{x}_1-\overline{x}_2\right)\pm z_\alpha\cdot\sqrt{\frac{n_1.s_1^2+n_2.s_2^2}{n_1+n_2-2}}\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

B) Muestras pequeñas ($n_1+n_2-2 \le 50$) :

Función de decisión :	Región de aceptación (bilateral) :
$t_{n_1+n_2-2,\alpha}=\dfrac{\left(\overline{x}_1-\overline{x}_2\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{n_1.s_1^2+n_2.s_2^2}{n_1+n_2-2}}\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$	$\left(\mu_1-\mu_2\right)\pm t_{n_1+n_2-2,\alpha}\cdot\sqrt{\frac{n_1.s_1^2+n_2.s_2^2}{n_1+n_2-2}}\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$
	Intervalo de confianza : $\left(\overline{x}_1-\overline{x}_2\right)\pm t_{n_1+n_2-2,\alpha}\cdot\sqrt{\frac{n_1.s_1^2+n_2.s_2^2}{n_1+n_2-2}}\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

3.3. Muestras independientes y varianzas de las poblaciones desconocidas y distintas :

A) Muestras pequeñas pero de igual tamaño $n_1 = n_2 = n$:

Función de decisión :	Región de aceptación (bilateral) :
$t_{2n-2,\alpha}=\dfrac{\left(\overline{x}_1-\overline{x}_2\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{s_1^2+s_2^2}{n-1}}}$	$\left(\mu_1-\mu_2\right)\pm t_{2n-2,\alpha}\cdot\sqrt{\frac{s_1^2+s_2^2}{n-1}}$
	Intervalo de confianza : $\left(\overline{x}_1-\overline{x}_2\right)\pm t_{2n-2,\alpha}\cdot\sqrt{\frac{s_1^2+s_2^2}{n-1}}$

B) En otro caso, calcularemos previamente la expresión "*gl*" siguiente (grados de libertad) :

$$gl=\frac{\left[\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right]^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1+1}+\frac{\left(s_2^2/n_2\right)^2}{n_2+1}}-2$$

B.1) Muestras grandes (gl > 50) :

Función de decisión :	Región de aceptación (bilateral) :
$z_\alpha=\dfrac{\left(\overline{x}_1-\overline{x}_2\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{s_1^2}{n_1-1}+\frac{s_2^2}{n_2-1}}}$	$\left(\mu_1-\mu_2\right)\pm z_\alpha\cdot\sqrt{\frac{s_1^2}{n_1-1}+\frac{s_2^2}{n_2-1}}$
	Intervalo de confianza : $\left(\overline{x}_1-\overline{x}_2\right)\pm z_\alpha\cdot\sqrt{\frac{s_1^2}{n_1-1}+\frac{s_2^2}{n_2-1}}$

B.2) Muestras pequeñas (gl ≤ 50):

Función de decisión :	Región de aceptación (bilateral) :
$t_{gl,\alpha}=\dfrac{\left(\overline{x}_1-\overline{x}_2\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{s_1^2}{n_1-1}+\frac{s_2^2}{n_2-1}}}$	$\left(\mu_1-\mu_2\right)\pm t_{gl,\alpha}\cdot\sqrt{\frac{s_1^2}{n_1-1}+\frac{s_2^2}{n_2-1}}$
	Intervalo de confianza : $\left(\overline{x}_1-\overline{x}_2\right)\pm t_{gl,\alpha}\cdot\sqrt{\frac{s_1^2}{n_1-1}+\frac{s_2^2}{n_2-1}}$

NOTA :

Las expresiones $\frac{s_i'^2}{n_i}$ (en términos de cuasivarianzas) , pueden ser sustituidas por $\frac{s_i^2}{n_i - 1}$ (en función de las varianzas muestrales).

3.4. Muestras relacionadas y varianza de la diferencia σ_D^2 conocida :

Función de decisión :	Región de aceptación (bilateral) :	Intervalo de confianza :
$z_\alpha = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sigma_D / \sqrt{n}}$	$\left(\mu_1 - \mu_2\right) \pm z_\alpha . \sigma_D / \sqrt{n}$	$\left(\bar{x}_1 - \bar{x}_2\right) \pm z_\alpha . \sigma_D / \sqrt{n}$

3.5. Muestras relacionadas y varianza de la diferencia desconocida (*muestras grandes: n > 51*):

Siendo : $s_D^2 = \frac{\sum\left(D_i - \bar{D}\right)^2}{n}$

Función de decisión :	Región de aceptación (bilateral) :	Intervalo de confianza :
$z_\alpha = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{s_D / \sqrt{n-1}}$	$\left(\mu_1 - \mu_2\right) \pm z_\alpha . s_D / \sqrt{n-1}$	$\left(\bar{x}_1 - \bar{x}_2\right) \pm z_\alpha . s_D / \sqrt{n-1}$

3.6. Muestras relacionadas y varianza de la diferencia desconocida (*muestras pequeñas: n ≤ 51*):

Función de decisión :	Región de aceptación (bilateral) :	Intervalo de confianza :
$t_{n-1,\alpha} = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{s_D / \sqrt{n-1}}$	$\left(\mu_1 - \mu_2\right) \pm t_{n-1,\alpha} . s_D / \sqrt{n-1}$	$\left(\bar{x}_1 - \bar{x}_2\right) \pm t_{n-1,\alpha} . s_D / \sqrt{n-1}$

4. COCIENTE DE VARIANZAS (Hipótesis de igualdad [Cociente = 1])

A) Observaciones relacionadas :

Función de decisión :
$t_{n-2,\alpha} = \frac{(s_1'^2 - s_2'^2).\sqrt{n-2}}{2.s'_1.s'_2.\sqrt{1-r^2}} = \frac{(s_1^2 - s_2^2).\sqrt{n-2}}{2.s_1.s_2.\sqrt{1-r^2}}$

En la expresión anterior, **r** representa el coeficiente de correlación.

En este supuesto, no procede el cálculo de intervalos.

B) Observaciones independientes :

CASO GENERAL : Medias poblacionales desconocidas.

Función de decisión :	Intervalo de confianza :
$F_{n_1-1,n_2-1,\alpha} = \frac{s_1'^2}{s_2'^2}.\frac{\sigma_2^2}{\sigma_1^2} = \frac{\frac{n_1}{n_1-1}.s_1^2}{\frac{n_2}{n_2-1}.s_2^2}.\frac{\sigma_2^2}{\sigma_1^2}$	$\left(\frac{\frac{n_1}{n_1-1}.s_1^2}{\frac{n_2}{n_2-1}.s_2^2}.F_1 \, , \, \frac{\frac{n_1}{n_1-1}.s_1^2}{\frac{n_2}{n_2-1}.s_2^2}.F_2\right) = \left(\frac{s_1'^2}{s_2'^2}.F_1 \, , \, \frac{s_1'^2}{s_2'^2}.F_2\right)$
Región de aceptación (bilateral) : $\frac{s_1^2}{s_2^2} \text{ en } \left(\frac{\frac{n_2}{n_2-1}.\sigma_1^2}{\frac{n_1}{n_1-1}.\sigma_2^2}.F_1 \, , \, \frac{\frac{n_2}{n_2-1}.\sigma_1^2}{\frac{n_1}{n_1-1}.\sigma_2^2}.F_2\right)$ o bien: $\frac{s_1'^2}{s_2'^2} \text{ en } \left(\frac{\sigma_1^2}{\sigma_2^2}.F_1 \, , \, \frac{\sigma_1^2}{\sigma_2^2}.F_2\right)$	**NOTA :** En todos los casos, lo habitual es que el cociente de varianzas poblacionales se suprima, al ser igual a la unidad : $\frac{\sigma_1^2}{\sigma_2^2} = \frac{\sigma_2^2}{\sigma_1^2} = 1$

CASO PARTICULAR (poco frecuente) : Medias poblacionales conocidas.

Función de decisión : $F_{n_1-1,n_2-1,\alpha} = \frac{n_2}{n_1}.\frac{\sum\left(x_1 - \mu_1\right)^2}{\sum\left(x_2 - \mu_2\right)^2}.\frac{\sigma_2^2}{\sigma_1^2} = \frac{s_1^2 + \left(\bar{x}_1 - \mu_1\right)^2}{s_2^2 + \left(\bar{x}_2 - \mu_2\right)^2}.\frac{\sigma_2^2}{\sigma_1^2}$

5. PROPORCIÓN (valor de 0 a 1)

Muestras grandes (n > 51):

Función de decisión :	Región de aceptación (bilateral) :	Intervalo de confianza :
$z_\alpha = \dfrac{p-\pi}{\sqrt{\dfrac{\pi.(1-\pi)}{n}}}$	$\pi \pm z_\alpha \cdot \sqrt{\dfrac{\pi.(1-\pi)}{n}}$	$p \pm z_\alpha \cdot \sqrt{\dfrac{p.(1-p)}{n}}$

Muestras pequeñas (n ≤ 51):

Función de decisión :	Región de aceptación (bilateral) :	Intervalo de confianza :
$t_{n-1,\alpha} = \dfrac{p-\pi}{\sqrt{\dfrac{\pi.(1-\pi)}{n}}}$	$\pi \pm t_{n-1,\alpha} \cdot \sqrt{\dfrac{\pi.(1-\pi)}{n}}$	$p \pm t_{n-1,\alpha} \cdot \sqrt{\dfrac{p.(1-p)}{n}}$

6. DIFERENCIA DE PROPORCIONES

6.1. Muestras independientes, con $\pi_1 - \pi_2 = 0$:

Siendo

$$D_{decisión} = \sqrt{\left(\frac{n_1.\pi_1 + n_2.\pi_2}{n_1+n_2}\right).\left(1-\frac{n_1.\pi_1 + n_2.\pi_2}{n_1+n_2}\right).\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$D_{estimación} = \sqrt{\left(\frac{n_1.p_1 + n_2.p_2}{n_1+n_2}\right).\left(1-\frac{n_1.p_1 + n_2.p_2}{n_1+n_2}\right).\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Función de decisión :	Región de aceptación (bilateral) :	Intervalo de confianza :
$z_\alpha = \dfrac{p_1-p_2}{D_{estimación}}$	$(\pi_1-\pi_2) \pm z_\alpha . D_{decisión}$	$(p_1-p_2) \pm z_\alpha . D_{estimación}$

6.2. Muestras independientes, con $\pi_1 - \pi_2 \neq 0$:

Función de decisión :	Región de aceptación (bilateral) :
$z_\alpha = \dfrac{(p_1-p_2)-(\pi_1-\pi_2)}{\sqrt{\dfrac{p_1.(1-p_1)}{n_1}+\dfrac{p_2.(1-p_2)}{n_2}}}$	$(\pi_1-\pi_2) \pm z_\alpha \cdot \sqrt{\dfrac{\pi_1.(1-\pi_1)}{n_1}+\dfrac{\pi_2.(1-\pi_2)}{n_2}}$
	Intervalo de confianza : $(p_1-p_2) \pm z_\alpha \cdot \sqrt{\dfrac{p_1.(1-p_1)}{n_1}+\dfrac{p_2.(1-p_2)}{n_2}}$

NOTA :

En los dos supuestos anteriores, si $n_1 + n_2 - 2 \leq 50$, sustituiremos el valor normal tabulado $\mathbf{z_\alpha}$ por el de **t** de Student con $\mathbf{n_1 + n_2 - 2}$ grados de libertad.

6.3. Muestras relacionadas, con $\pi_1 - \pi_2 = 0$:

		Situación 2ª		
		1	**0**	
Situación 1ª	**0**	**A**	**B**	**A+B**
	1	**C**	**D**	**C+D**
		A+C	**B+D**	**n**

NOTA : A, B, C y D representan frecuencias.

Función de decisión :	Intervalo de confianza :
$z_\alpha = \dfrac{D-A}{\sqrt{A+D}}$	$D \pm z_\alpha . \sqrt{(D+A).0'5.0'5}$

EJERCICIOS RESUELTOS (Estimación)

1

Con un nivel de confianza del 95%, estimar la media poblacional a partir de 15 observaciones muestrales en las que se calculó : media = 8'5 ; varianza = 3 .

Con los mismos resultados muestrales, estime el valor de la desviación típica de la población con un margen de error del 1%.

1º Estimación de la media de la población. Varianza poblacional desconocida y muestra pequeña.

Intervalo de confianza :

$$\bar{\mathbf{x}} \pm \mathbf{t}_{\mathbf{n-1,\alpha}} \cdot \frac{\mathbf{s'}}{\sqrt{\mathbf{n}}} \qquad \text{con } \mathbf{s'}^2 = \frac{\mathbf{n}}{\mathbf{n-1}} . \mathbf{s}^2 = \text{covarianza}$$

o bien, sustituyendo **s'** :

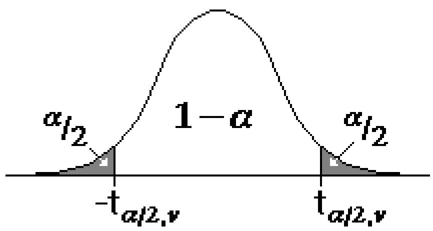
$$\bar{\mathbf{x}} \pm \mathbf{t}_{\mathbf{n-1,\alpha}} \cdot \frac{\mathbf{s}}{\sqrt{\mathbf{n-1}}}$$

Valores tabulados de t de Student para :

$1-\alpha = 0'95$ ($\alpha = 0'05$)

$\nu = n-1 = 14$ grados de libertad

$t_{14,\alpha} = \pm 2'144776$



El intervalo de confianza para la estimación de la media es :

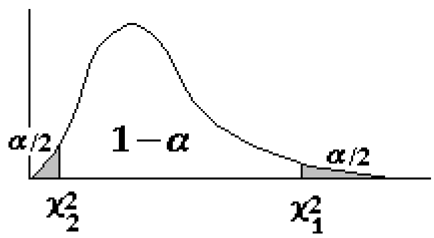
$$8'5 \pm 2'144776 . \frac{\sqrt{3}}{\sqrt{15-1}} \quad \Rightarrow \quad 8'5 \pm 0'993 \quad \Rightarrow \quad (7'507 \ , 9'493)$$

2º Estimación de la desviación típica poblacional.

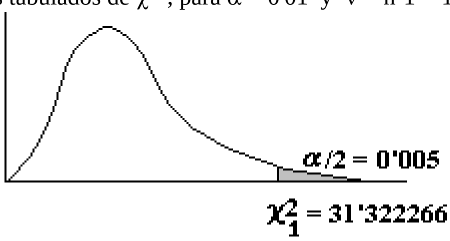
Realizamos el proceso para la varianza de la población. Una vez calculado el intervalo de confianza, la raíz cuadrada de sus extremos proporcionará el intervalo correspondiente a la desviación típica.

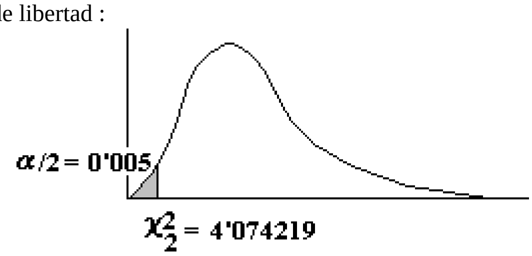
Intervalo de confianza :

$$\left(\frac{n.s^2}{\chi_1^2} , \frac{n.s^2}{\chi_2^2} \right)$$



Valores tabulados de χ^2 , para $\alpha = 0'01$ y $\nu = n-1 = 14$ grados de libertad :





Luego el intervalo de confianza para la varianza es : $\left(\frac{15 \ .3}{31'322266} , \frac{15 \ .3}{4'074219} \right) = \left(1'4367 \ , 11'0451 \right)$

y el correspondiente a la desviación típica : $\left(\sqrt{1'4367} \ , \sqrt{11'0451} \right) = \left(1'1986 \ , 3'3234 \right)$

2

Observadas las calificaciones en las asignaturas de Filosofía y Lengua se aprecia una notable diferencia entre ellas. Con el fin de analizarla se seleccionan 20 alumnos de Filosofía y otros tantos de Lengua, obteniendo :

Filosofía :	**Media = 6**	**Varianza = 2**
Lengua :	**Media = 5**	**Varianza = 1'5**

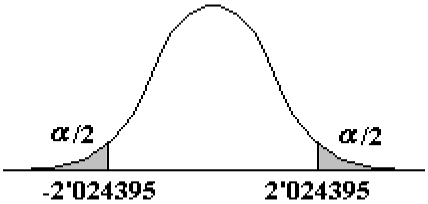
Estime la diferencia media de las calificaciones con un margen de error del 5%.

Estimación de una diferencia de medias con varianzas poblacionales desconocidas y muestras pequeñas de igual tamaño.

Intervalo de confianza :

$$\left(\overline{x}_1 - \overline{x}_2\right) \pm t_{2n-2,\alpha} \cdot \sqrt{\frac{s_1^2 + s_2^2}{n-1}} \qquad \text{con } \mathbf{n = n_1 = n_2}$$

Valores tabulados de la t de Student para α = 0'05 y ν = 2n-2 = 38 grados de libertad :



$t_{2n-2,\alpha} = t_{38,0'05} = \pm 2'024395$

El intervalo pedido es :

$$(6-5) \pm 2'024395. \sqrt{\frac{2+1'5}{20-1}} \quad \Rightarrow \quad 1 \pm 0'869 \quad \Rightarrow \quad \left(0'131\,,\,1'869\right)$$

La diferencia entre los promedios de calificaciones estará comprendida entre 0'131 y 1'869 puntos.

3

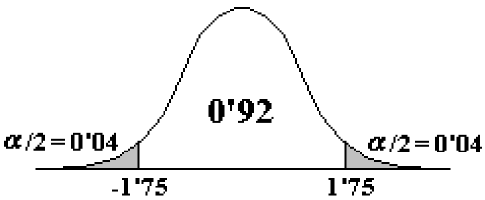
Para analizar la eficacia de la aplicación de un tratamiento , se someten al mismo a 64 pacientes. Finalizado el período de aplicación se observó que remitió la enfermedad en 50 casos.
Con un nivel de confianza del 92%, estime el porcentaje de efectividad del tratamiento objeto de estudio.

Estimación de una proporción.

Intervalo de confianza :

$$\mathbf{p \pm z_\alpha \cdot \sqrt{\frac{p.(1-p)}{n}}}$$

Valor tabulado en N(0,1) para α = 0'08 (p = 1-α = 0'92) :



$z_\alpha = \pm 1'75$

Siendo la proporción muestral p = 50 / 64 = 0'78125 , el intervalo de confianza es :

$$0'78125 \pm 1'75. \sqrt{\frac{0'78125.(1-0'78125)}{64}} \Rightarrow 0'78125 \pm 0'09043 \Rightarrow \left(0'69082, 0'87168\right)$$

Con un margen de error del 8% podemos afirmar que el tratamiento será efectivo entre el 69'082% y el 87'168% de los casos.

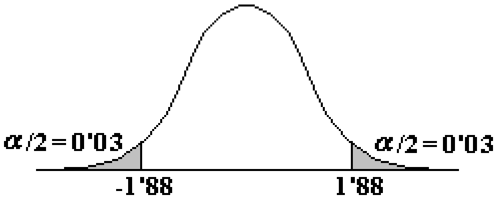
4

Por experiencias anteriores se sabe que las estaturas de los soldados tienen una varianza de 64 cm.
a) Con un margen de error del 6%, estime la estatura media partiendo de un grupo de 100 soldados, tomados al azar, sabiendo que proporcionó un promedio de 164 cm.
b) Manteniendo el mismo intervalo de confianza calculado en el apartado anterior, ¿ qué tamaño muestral debe fijarse para que el margen de error pase a ser α = 0'02 ?.

a) Estimación de la media con varianza poblacional conocida.

Intervalo de confianza :

$$\mathbf{\overline{x} \pm z_\alpha \cdot \frac{\sigma}{\sqrt{n}}}$$



Intervalo de confianza :

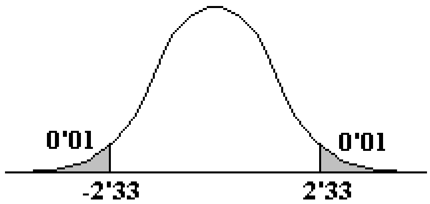
$$164 \pm 1'88.\frac{\sqrt{64}}{\sqrt{100}}$$

$$164 \pm 1'504$$

(162'496 , 165'504)

b)

Para α = 0'02 ; $z_\alpha = \pm$ 2'33



Manteniendo el intervalo anterior, tomemos uno de sus extremos, por ejemplo el derecho :

$$165'504 = \bar{x} + z_\alpha.\frac{\sigma}{\sqrt{n}} = 164 + 2'33.\frac{\sqrt{64}}{\sqrt{n}}$$

De aquí resulta : $\sqrt{n} = \dfrac{2'33.\sqrt{64}}{165'504 - 164} = 12'3936 \quad \Rightarrow \quad n = 153'60 \cong 154$

Para reducir a 0'02 el margen de error, debemos tomar una muestra de 154 soldados.

5

Un estudio sobre la proporción de fumadores estableció que entre el personal de un Hospital sólo fumaban entre el 20% y el 35%.
a) ¿ Cuál fue el porcentaje muestral observado ?.
b) Si se realizó el estudio partiendo de una muestra de 40 individuos, ¿ con qué margen de error se trabajó ?.
c) Si el análisis se efectuó con un nivel de confianza del 95%, ¿ cuántos individuos integraron la muestra seleccionada ?.

Estimación de una proporción.

Intervalo de confianza :

$$p \pm z_\alpha.\sqrt{\frac{p.(1-p)}{n}}$$

a)

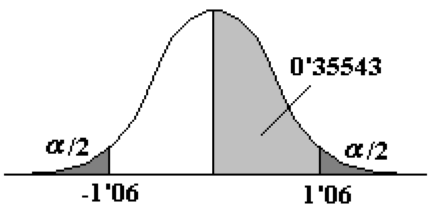
El valor central del intervalo es siempre la proporción muestral. En consecuencia, al ser el intervalo de confianza (0'20 , 0'35), resulta : **p = proporción muestral = 0'275**.

b)

Tomando uno de los extremos del intervalo, por ejemplo el izquierdo :

$$0'20 = p - z_\alpha.\sqrt{\frac{p.(1-p)}{n}} = 0'275 - z_\alpha.\sqrt{\frac{0'275.(1-0'275)}{40}} = 0'275 - z_\alpha.0'0706$$

De aquí : $z_\alpha = \dfrac{0'275 - 0'20}{0'0706} = 1'06232 \cong 1'06$



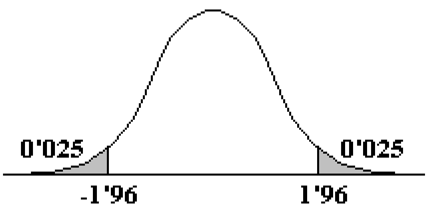
Luego : $\alpha/2$ = 0'5 - 0'35543 = 0'14457

α = 2 . 0'14457 = **0'28914**

Es decir, el estudio se realizó con un margen de error del 28'914%.

c)

Para un margen de confianza del 95% (α = 0'05), se obtiene un valor tabulado : $z_\alpha = \pm$ 1'96.



Tomando uno de los extremos del intervalo (sea ahora el derecho)

$$0'35 = p + z_\alpha.\sqrt{\frac{p.(1-p)}{n}} = 0'275 + 1'96.\sqrt{\frac{0'275.(1-0'275)}{n}}$$

podemos obtener el tamaño muestral

$$0'35 - 0'275 = 1'96 . \frac{0'4465}{\sqrt{n}} \quad \Rightarrow \quad \sqrt{n} = \frac{1'96 . 0'4465}{0'35 - 0'20} = 11'6685$$

El tamaño de la muestra fue de **n =** $11'6685^2 = 136'154 \cong$ **137 individuos**

6

En una muestra de 13 elementos, obtenidos aleatoriamente de una población normal e infinita, sabemos que los límites del intervalo confidencial de estimación de la media poblacional, con un nivel de riesgo del 5% son, respectivamente, 1'28 y 18'72. Calcule :

a) la media de la muestra

b) la desviación típica insesgada de la muestra.

a) La media muestral siempre ocupa el valor central o medio del intervalo de confianza. Así :

$$\bar{x} = \frac{18'72 - 1'28}{2} = 8'72$$

b)

En la estimación de la media poblacional con varianza desconocida y siendo la muestra de tamaño pequeño, el intervalo de confianza es :

$$\left(\bar{x} - t_{n-1,\alpha/2} . \frac{s'}{\sqrt{n}} \ , \ \bar{x} + t_{n-1,\alpha/2} . \frac{s'}{\sqrt{n}} \right) = (1'28 \ , 18'72)$$

Las tablas de la t de Student proporcionan el valor : $t_{n-1,\alpha/2} = t_{12,0'05/2} = 2'178792$

Tomando uno cualquiera de los extremos del intervalo de confianza obtenemos :

$$\bar{x} + t_{n-1,\alpha/2} . \frac{s'}{\sqrt{n}} = 8'72 + 2'178792 . \frac{s'}{\sqrt{13}} = 18'72 \quad \Rightarrow \quad 2'178792 . \frac{s'}{\sqrt{13}} = 10 \quad \Rightarrow \quad s' = \frac{10 . \sqrt{13}}{2'178792} = 16'548$$

7

Partiendo de una muestra de 113 elementos, se obtuvo 11'856 como límite superior del intervalo de estimación de la varianza poblacional, con un margen de error del 5%.
Calcule la varianza insesgada de la muestra.

En la estimación de una varianza, el intervalo de confianza es :

$$\left(\frac{n.s^2}{\chi_1^2} , \frac{n.s^2}{\chi_2^2} \right) \quad \text{o bien} \quad \left(\frac{(n-1).s'^2}{\chi_1^2} , \frac{(n-1).s'^2}{\chi_2^2} \right)$$

Determinemos los valores tabulados de χ^2. En este caso, al ser el número de grados de libertad superior a 50 (n-1 = 112), los calculamos mediante la tabla de la distribución normal.

Los valores z correspondientes a $\alpha/2 = 0'025$ son : $z = \pm 1'96$

De aquí :

para z = 1'96 : $\chi^2 = \frac{1}{2} . \left(z + \sqrt{2.\upsilon - 1} \right)^2 = \frac{1}{2} . \left(1'96 + \sqrt{2 \ . \ 112 - 1} \right)^2 = 142'69$

para z = -1'96 : $\chi^2 = \frac{1}{2} . \left(z + \sqrt{2.\upsilon - 1} \right)^2 = \frac{1}{2} . \left(-1'96 + \sqrt{2 \ . \ 112 - 1} \right)^2 = 84'15$

Sustituyendo en el intervalo de confianza obtendremos la varianza insesgada (cuasivarianza) s'^2 :

$$\left(\frac{(n-1).s'^2}{\chi_1^2} , \frac{(n-1).s'^2}{\chi_2^2} \right) = \left(\frac{112.s'^2}{142'69} , \frac{112.s'^2}{84'15} \right) \quad \Rightarrow \quad \frac{112.s'^2}{84'15} = 11'856 \quad \Rightarrow \quad s'^2 = \frac{11'856 . 84'15}{112} = 8'908$$

EJERCICIOS PROPUESTOS (Estimación)

1

Una muestra aleatoria de 40 mujeres de una población dio una estatura media de 156 cm. con desviación típica 3 cm.. En otra muestra de 60 hombres se calculó que la media era de 160 cm. con varianza 25.
Con un nivel de confianza del 95%, estimar la diferencia de estaturas entre hombres y mujeres.

2

De 100 alumnos seleccionados aleatoriamente de un Centro Universitario, 55 eran mujeres. Con márgenes de error del 5% y del 1%, determine la proporción de mujeres del Centro.

3

Analizadas 16 muestras de sangre se determinó 64 mg./dl. para la varianza de las determinaciones de colesterol. Con un nivel de confianza del 95%, estime la desviación típica de las determinaciones de colesterol en sangre.

4

En la medición de longitudes de tornillos producidos por una máquina se calculó una desviación típica de 0'05 mm. . Seleccionada una muestra aleatoria de ellos dio un promedio de 21 mm.
Con un margen de error del 5%, ¿qué tamaño debe tener la muestra seleccionada para poder afirmar que la longitud media de los tornillos está comprendida entre 20'99 y 21'01 mm. ?.

5

Partiendo del estudio de una muestra de 90 individuos a los que se aplicó un test de agresividad, se obtuvo 25 puntos para su desviación típica, estimando que la puntuación media del test debería estar comprendida entre 64 y 80 puntos.
Calcule la media muestral del test y el margen de error con el que se realizó la estimación.

SOLUCIONES DE LOS EJERCICIOS PROPUESTOS

1

Estimación de diferencias de medias con varianzas poblacionales desconocidas (muestras grandes).
(2'414 , 5'586)

2

Estimación de una proporción. p = 0'55
a) Para α = 0'05 : (0'45 , 0'65)
b) Para α = 0'01 : (0'42 , 0'68)

3

Estimación de la varianza de la población : (37'236 , 163'578)
Estimación de la desviación típica : (6'102 , 12'790)

4

Estimación de la media con varianza poblacional conocida. n = 96

5

Estimación de la media para varianza poblacional desconocida y muestra grande.
Media muestral = 72 ; z_α = 3'02 ; α = 0'00254

EJERCICIOS RESUELTOS (Decisión)

1

En un Hospital se determinó que el promedio de colesterol en sangre no se ajusta al estándar establecido de 180 mg./dl.
Para comprobarlo, seleccionamos aleatoriamente a 15 individuos, obteniendo las determinaciones siguientes :

196	**175**	**190**	**184**	**179**	**187**	**203**	**168**	**190**	**182**
178	**183**	**198**	**200**	**175**					

¿ Apoyan estos datos muestrales la sospecha del Centro hospitalario, con un nivel de significación del 5% ?.

La resolución de problemas de decisión, puede realizarse de dos formas distintas :

1. Utilizando valores tipificados, a través de la función de decisión. En este caso, la hipótesis nula será aceptada si el valor experimental de la función de decisión se encuentra dentro de la región de aceptación tipificada, definida por los valores que directamente proporcionan las tablas.
2. Mediante los valores reales de la característica objeto de estudio. Construiremos la región de aceptación, aceptando la hipótesis nula si el valor muestral del estadístico a contrastar se encuentra dentro de dicha región.

Nos encontramos aquí con un contraste de la media poblacional (varianza de la población desconocida ; muestra pequeña).

Hipótesis nula y alternativa :

$H_0 : \ \mu = 180$
$H_1 : \ \mu \neq 180$

Parámetros muestrales :

Media = $\overline{\mathbf{x}}$ = 185'87
Desviación típica = **s** = 9'8986
Desviación estimada (raíz de la cuasivarianza) = **s'** = 10'246

$$\overline{x} = \frac{\sum x_i}{n}$$
$$s = \sqrt{\frac{\sum x_i^2}{n} - \overline{x}^2}$$
$$s'^2 = \frac{n}{n-1}.s^2$$

Método 1º : Utilizando la función de decisión.

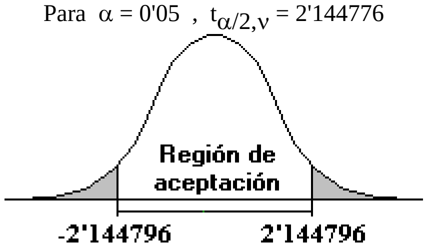
Función de decisión :

$$\mathbf{t} = \frac{\overline{\mathbf{x}} - \mu}{\mathbf{s}' / \sqrt{\mathbf{n}}} = \frac{\mathbf{185'87 - 180}}{\mathbf{10'246} / \sqrt{\mathbf{15}}} = \mathbf{2'2189}$$

se distribuye según una t de Student con **n-1** = 14 grados de libertad.

Decisión :

Niveles críticos de decisión (tipificados) :
Para α = 0'05 , $t_{\alpha/2,\nu}$ = 2'144776



Al no estar t = 2'2198 comprendido entre los dos niveles críticos de decisión -2'144796 y 2'144796 (no pertenece a la región de aceptación tipificada), rechazamos la hipótesis nula.

Aceptamos con ello la sospecha establecida en el Centro hospitalario.

Método 2º : Utilizando la región de aceptación.

Región de aceptación : (Contraste bilateral)

Con t de Student de **n-1** = 14 grados de libertad, encontramos como valor tabulado para α = 0'05 :
$t_{\alpha/2,\nu}$ = 2'144776

$$\mu \pm t_{n-1,\alpha} . \frac{s'}{\sqrt{n}} = \left(180 - 2'144776.\frac{10'246}{\sqrt{15}} , 180 + 2'144776.\frac{10'246}{\sqrt{15}}\right) = (174'326 , 185'741)$$

Decisión :

Como la media muestral (185’87) no pertenece a la región de aceptación, rechazamos la hipótesis nula, aceptando así la sospecha establecida en el Centro hospitalario.

2

Un psicólogo educacional considera que el número de alumnos que abandonan los estudios es inferior al 15% que se establece por las autoridades educativas. Para comprobar la certeza de su creencia, selecciona al azar un grupo de 500 alumnos resultando que sólo 59 dejaron los estudios.
¿ Podemos aceptar la hipótesis planteada por el psicólogo, con un nivel de significación del 5% ?.

Contraste paramétrico de la proporción poblacional.

Hipótesis nula y alternativa :

$H_0 : \ \pi \geq 0'15$
$H_1 : \ \pi < 0'15$

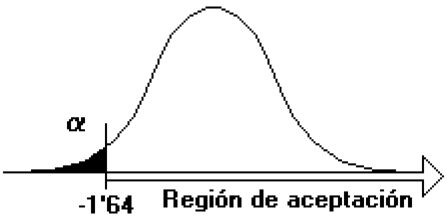
Observe que se ha construido la hipótesis alternativa en base a la hipótesis experimental. El rechazo de la hipótesis nula permitirá aceptar lo planteado por el psicólogo.
No sería una incorrección plantear la hipótesis nula en los términos $H_0 : \pi = 0'15$, manteniendo la alternativa anterior. Es precisamente la hipótesis alternativa la que se debe fijar con sumo cuidado, ya que ella determina la región de rechazo del contraste y el tipo : bilateral, unilateral derecho o unilateral izquierdo.

Función de decisión :
Proporción muestral : $p = 59/500 = 0'118$
El estadístico de contraste se distribuye normalmente, siendo :

$$\mathbf{z = \frac{p-\pi}{\sqrt{\frac{\pi.(1-\pi)}{n}}} = \frac{0'118-0'15}{\sqrt{\frac{0'15.0'85}{500}}} = -2'0039}$$

Criterio de decisión :

Nivel crítico de decisión :
Para $\alpha = 0'05$, $z_\alpha = -1'64$



La formulación de las hipótesis hace que rechacemos la hipótesis nula si $\mathbf{z < z_\alpha}$.

Como **-2'0039 < -1'64** , rechazamos la hipótesis nula.

Aceptamos así lo supuesto por el psicólogo.

Método basado en la región de aceptación.

La hipótesis alternativa (H_1 : $\pi < 0'15$) indica que nos encontramos ante un contraste unilateral. La región de aceptación será desde un cierto valor en adelante : (A , +∞).

$$\pi \pm z_\alpha . \sqrt{\frac{\pi.(1-\pi)}{n}} \quad \rightarrow \quad \left(\pi - z_\alpha . \sqrt{\frac{\pi.(1-\pi)}{n}} \ , +\infty\right) = \left(0'15 - 1'64. \sqrt{\frac{0'15.0'85}{500}} \ , +\infty\right) = \left(0'1238 \ , +\infty\right)$$

Como la proporción muestral (0’118) se encuentra dentro de la región de aceptación, admitimos lo establecido en la hipótesis nula.

3

Durante los últimos años se ha establecido en 1'1 la desviación típica de las calificaciones en las pruebas de acceso a la Universidad.
Ante la sospecha de que en la actualidad ha aumentado la dispersión de las calificaciones, se seleccionan las correspondientes a 25 alumnos presentados en la última convocatoria, resultando como desviación típica 1'64.
Con un nivel de significación del 1%, ¿ podemos concluir que aumentó la dispersión de las calificaciones de las pruebas de acceso ?.

Contraste paramétrico de la varianza poblacional.

Hipótesis nula y alternativa :

$H_0 : \ \sigma^2 \leq 1'21 \quad (1'1^2)$
$H_1 : \ \sigma^2 > 1'21$

Construimos la hipótesis alternativa en base a la hipótesis experimental, expresada siempre en términos de varianzas (nunca desviaciones típicas).

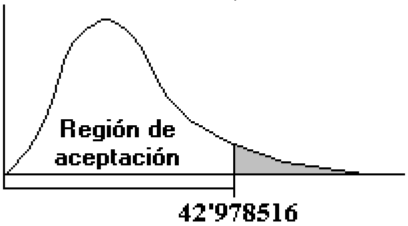
Función de decisión : El estadístico de contraste se distribuye según una χ^2 con $\nu = n-1 = 24$ grados de libertad.

$$\chi^2 = \frac{n.s^2}{\sigma^2} = \frac{25.1'64^2}{1'21} = 55'57025$$

Criterio de decisión :

Nivel crítico de decisión :

Para $\alpha = 0'01$, $\chi^2_{\alpha,\nu} = 42'978516$



Siendo $\chi^2 > \chi^2_{\alpha,\nu}$ (55'57025 > 42'978516), rechazamos la hipótesis nula.

La dispersión de las calificaciones en la prueba de acceso se ha incrementado.

<u>Método basado en la región de aceptación.</u>

La hipótesis alternativa (H_1 : $\sigma^2 > 1'21$) indica que nos encontramos ante un contraste unilateral. La región de aceptación será desde un cierto valor hacia atrás : ($-\infty$, A).

Para $\alpha = 0'01$, las tablas proporcionan : $\chi^2_{\alpha,\nu} = 42'978516$

$$\left(\frac{\sigma^2 . \chi_1^2}{n}, \frac{\sigma^2 . \chi_2^2}{n}\right) \rightarrow \left(-\infty, \frac{\sigma^2 . \chi^2}{n}\right) = \left(-\infty, \frac{1'21 . 42'978516}{25}\right) = \left(-\infty , 2'08\right)$$

Puesto que la varianza muestral ($1'64^2 = 2'6896$) queda fuera de la región de aceptación, rechazamos la hipótesis nula.

Recordemos que el problema trataba de desviaciones típicas. Este segundo método nos permite, en este caso, definir la región de aceptación para dicho estadístico (desviaciones típicas) :

Región de aceptación de la desviación típica : $\left(-\infty , \sqrt{2'08}\right) = \left(-\infty , 1'44\right)$

Puesto que la desviación típica muestral 1'64 no pertenece al intervalo ($-\infty$, 1'44) , rechazamos la hipótesis nula.

4

Las calificaciones de 18 alumnos en Psicología Matemática (PM) y en Psicología General (PG) fueron las siguientes :

Alumno	1	2	3	4	5	6	7	8	9
PM	6'3	5'1	3'5	2	8	6'4	7	5'1	5
PG	7	6'2	6'5	4'8	9'5	6	5'8	4	4'5

Alumno	10	11	12	13	14	15	16	17	18
PM	8'6	3	2'5	4	5'8	6	8'7	6	3
PG	9	3'8	5	6'5	7	5'2	9	5'8	4'5

Con una significación $\alpha = 0'05$, ¿ podemos afirmar que las calificaciones en ambas asignaturas están igualmente dispersas ?.

Comparación de varianzas para observaciones relacionadas.

Hipótesis :

$H_0 : \sigma_1^2 = \sigma_2^2$
$H_1 : \sigma_1^2 \neq \sigma_2^2$

Función de decisión :

El siguiente estadístico de contraste se distribuye según una t de Student con $\nu = n-2 = 16$ grados de libertad :

$$t = \frac{(s_1^2 - s_2^2).\sqrt{n-2}}{2.s_1.s_2.\sqrt{1-r^2}}$$

Determinando los parámetros muestrales necesarios, obtenemos :

hPsicología Matemática : $\bar{x}_1 = 5'333$; $s_1 = 1'9785$

Psicología General : $\bar{x}_2 = 6'117$; $s_2 = 1'6507$

Covarianza : $s_{12} = 2'4708$

Coeficiente de correlación de Pearson : $r = \frac{s_{12}}{s_1 . s_2} = \frac{2'4708}{1'9785.1'6507} = 0'7565$

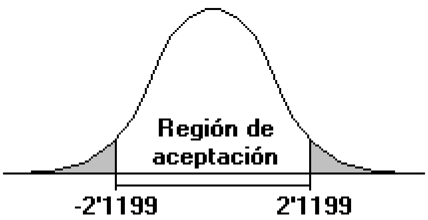
El estadístico de contraste toma el valor :

$$t = \frac{(s_1^2 - s_2^2).\sqrt{n-2}}{2.s_1.s_2.\sqrt{1-r^2}} = \frac{(1'9785^2 - 1'6507^2).\sqrt{18-2}}{2 \,.\, 1'9785 \,.\, 1'6507.\sqrt{1-0'7565^2}} = 1'1140$$

Criterio de decisión :

Nivel crítico de decisión :

Para $\alpha = 0'05$, $t_{\alpha/2,\nu} = 2'119900$



Al ser $-2'1199 < t < 2'1199$, aceptamos la hipótesis nula.

Las calificaciones en Psicología Matemática y en Psicología General tienen la misma variabilidad.

NOTA :

Ciertos autores establecen, en la función de decisión, cuasivarianzas (estimación insesgada de la varianza) en lugar de varianzas. Ambas expresiones son correctas, como comprobamos a continuación :

$$t = \frac{(s'^2_1 - s'^2_2).\sqrt{n-2}}{2.s'_1.s'_2.\sqrt{1-r^2}} = \frac{(\frac{n}{n-1}.s_1^2 - \frac{n}{n-1}s_2^2).\sqrt{n-2}}{2.\sqrt{\frac{n}{n-1}}.s_1.\sqrt{\frac{n}{n-1}}.s_2.\sqrt{1-r^2}} = \frac{\cancel{\frac{n}{n-1}}.(s_1^2 - s_2^2).\sqrt{n-2}}{\cancel{\frac{n}{n-1}}.2.s_1.s_2.\sqrt{1-r^2}} = \frac{(s_1^2 - s_2^2).\sqrt{n-2}}{2.s_1.s_2.\sqrt{1-r^2}}$$

5

Una academia de conductores desea experimentar sobre la eficacia de un nuevo método de enseñanza basado en el empleo de simuladores. Toma para ello al azar a 28 alumnos utilizando con 16 de ellos el método en experimentación y , con los demás, el tradicional.

Finalizado el curso de formación de conductores, los exámenes dan como resultados :

Alumnos que siguieron el método basado en simuladores :

Media = 6'8 Varianza = 1'4

Alumnos que siguieron el método tradicional :

Media = 5'8 Varianza = 2

¿ Qué podemos concluir sobre la eficacia del nuevo método ?.

NOTA : Realice el correspondiente contraste con niveles de significación del 10%, 5% y 1%.

Contraste de comparación de medias (diferencia de medias).

Consideración importante :

De forma premeditada se ha omitido en el enunciado todo tipo de indicación sobre las varianzas poblacionales. Como quiera que el estudio relativo a la comparación de dos medias, es distinto según sean iguales o distintas tales varianzas, deberíamos comenzar nuestro problema contrastando tal hipótesis (igualdad de varianzas poblacionales).

En función de su aceptación o rechazo, emplearemos el método que corresponda.

No lo hacemos en este caso, resolviendo el problema en las dos situaciones posibles.

Hipótesis :

$H_0 :\ \mu_1 = \mu_2 \qquad (\mu_1 - \mu_2 = 0)$

$H_1 :\ \mu_1 > \mu_2 \qquad (\mu_1 - \mu_2 > 0)$

1º.- Varianzas de las poblaciones desconocidas y las muestras pequeñas de distinto tamaño, estableciendo el supuesto previo de igualdad de varianzas poblacionales ($\sigma_1^2 = \sigma_2^2$).

Función de decisión :

El siguiente estadístico de contraste se distribuye según una t de Student con $\nu = n_1+n_2-2 = 26$ grados de libertad.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{n_1.s_1^2 + n_2.s_2^2}{n_1 + n_2 - 2}}.\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Los valores muestrales proporcionan : $t=\dfrac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{n_1.s_1^2+n_2.s_2^2}{n_1+n_2-2}}.\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}=\dfrac{6'8-5'8}{\sqrt{\dfrac{16.1'4+12.2}{26}}.\sqrt{\dfrac{1}{16}+\dfrac{1}{12}}}=1'9602$

El planteamiento de las hipótesis nos llevará a rechazar la hipótesis nula si el valor anterior es superior al tabulado.

α = 0'10	$t_{26,0'10}$ = 1'314972 < 1'9602	Se rechaza H_0. El método con simuladores es más eficaz que el tradicional
α = 0'05	$t_{26,0'05}$ = 1'705617 < 1'9602	Se rechaza H_0. El método con simuladores es más eficaz que el tradicional
α = 0'01	$t_{26,0'01}$ = 2'478629 > 1'9602	Se acepta H_0. El método con simuladores no puede asegurarse que sea significativamente más eficaz que el tradicional.

La disminución del nivel de significación α (probabilidad de cometer un error de tipo I) aumenta la región de aceptación de la hipótesis nula. Es por ello más fácil su aceptación, pero con una inferior significación del contraste.

2º.- Varianzas de las poblaciones desconocidas y las muestras pequeñas de distinto tamaño, estableciendo el supuesto previo de que las varianzas poblacionales son distintas ($\sigma_1^2 \neq \sigma_2^2$).

Comenzamos calculando la expresión que determina los grados de libertad :

$$gl=\frac{\left[\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}\right]^2}{\dfrac{\left(s_1^2/n_1\right)^2}{n_1+1}+\dfrac{\left(s_2^2/n_2\right)^2}{n_2+1}}-2=\frac{\left[\dfrac{1'4}{16}+\dfrac{2}{12}\right]^2}{\dfrac{\left(1'4/16\right)^2}{17}+\dfrac{\left(2/12\right)^2}{13}}-2=22'97\approx 23$$

Función de decisión :

Siendo los grados de libertad calculados inferiores o iguales a 30, el siguiente estadístico de contraste se distribuye según una t de Student con ν = gl = 23 grados de libertad.

$$t_{gl,\alpha}=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{s_1^2}{n_1-1}+\dfrac{s_2^2}{n_2-1}}}$$

Los valores muestrales proporcionan :

$$t_{gl,\alpha}=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{s_1^2}{n_1-1}+\dfrac{s_2^2}{n_2-1}}}=\frac{(6'8-5'8)-(0)}{\sqrt{\dfrac{1'4}{15}+\dfrac{2}{11}}}=1'9064$$

El planteamiento de las hipótesis nos llevará a rechazar la hipótesis nula si el valor anterior es superior al tabulado.

α = 0'10	$t_{23,0'10}$ = 1'319460 < 1'9604	Se rechaza H_0.
α = 0'05	$t_{23,0'05}$ = 1'713871 < 1'9604	Se rechaza H_0.
α = 0'01	$t_{23,0'01}$ = 2'499865 > 1'9604	Se acepta H_0.

6

El Ayuntamiento de una ciudad realiza una campaña de información sobre la labor desarrollada en los últimos años. Con el fin de analizar la eficacia de la campaña, seleccionamos al azar a 44 ciudadanos, preguntando su opinión sobre la gestión de los munícipes , al inicio y una vez concluida la misma.
Los resultados de la consulta fueron los siguientes :

		Después	
		Favor	**Contra**
Antes	**Favor**	**4**	**8**
	Contra	**27**	**5**

Con los niveles de significación del 5% y del 1%, ¿ podemos afirmar que ha cambiado la opinión de los ciudadanos ?.

Comparación de proporciones (diferencia de proporciones). Muestras relacionadas.

Hipótesis :

H_0 : $\pi_1 = \pi_2$ $\quad$ ($\pi_1 - \pi_2 = 0$)
H_1 : $\pi_1 \neq \pi_2$ $\quad$ ($\pi_1 - \pi_2 \neq 0$)

Función de decisión :

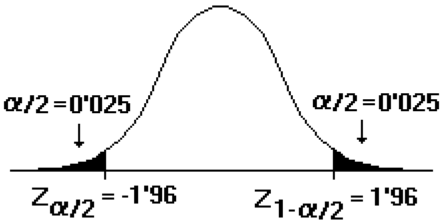
El estadístico $\mathbf{z = \dfrac{D - A}{\sqrt{A + D}}}$ se distribuye según una N(0,1).

		(2) Después	
		Contra	**Favor**
(1) Antes	**Favor**	**A = 8**	**B = 4**
	Contra	**C = 5**	**D = 27**

De aquí obtenemos : $\mathbf{z = \dfrac{D - A}{\sqrt{A + D}} = \dfrac{27 - 8}{\sqrt{8 + 27}} = 2'5355}$

Criterio de decisión :

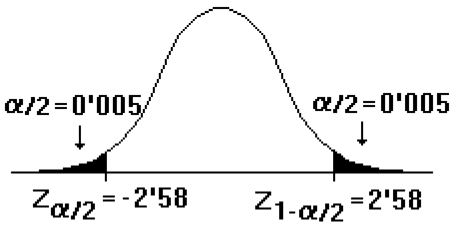
Al tratarse de un contraste bilateral, rechazamos la hipótesis nula si el estadístico de contraste anterior queda fuera del intervalo determinado por los niveles críticos $z_{\alpha/2}$ y $z_{1-\alpha/2}$.



Para α **= 0'05** :

Siendo **2'58 > 1'96** , rechazamos la hipótesis nula.

Podemos afirmar que existe una influencia significativa de la campaña en la opinión de los ciudadanos.



Para α **= 0'01** :

Siendo **-2'58 < 2'5355 < 2'58** , aceptamos la hipótesis nula.

No existen diferencias significativas para afirmar que exista influencia de la campaña en la opinión de los ciudadanos.

7

La aplicación de un test en el colectivo de profesionales de la enseñanza tiene, por experiencias anteriores, una puntuación media 55 con varianza 121. Un psicólogo educativo considera que en la actualidad el promedio se incrementó, pasando a ser de 60 puntos.

Para contrastar la hipótesis planteada por el psicólogo, se somete al test a 50 individuos en los que se obtiene una puntuación muestral media 58.

a) Con un nivel de significación α = 0'01, ¿ puede aceptarse el planteamiento del psicólogo ?.

b) Para el nivel α anterior, determine la probabilidad del error de tipo II.

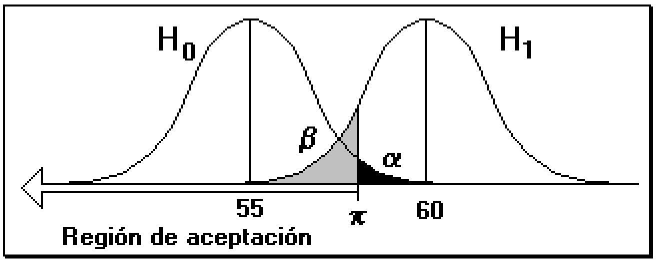
c) ¿ Qué tamaño muestral debe utilizarse para incrementar la potencia del contraste en un 10% ?.

Contraste de la media poblacional. Varianza de la población conocida.

Hipótesis nula y alternativa :

$H_0 : \ \mu = 55$

$H_1 : \ \mu = 60$



Función de decisión : La varianza de la población se conoce (σ^2 =121 ; σ = 11).

El estadístico de contraste se distribuye normalmente , siendo : $\mathbf{z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}}$

a) $\mathbf{z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \dfrac{58 - 55}{11 / \sqrt{50}} = 1'93}$; Valor tabulado : $z_{\alpha} = z_{0'01} = 2'33$

Aceptamos la hipótesis nula al ser 1'93 < 2'33 ($z < z_{\alpha}$), rechazando lo supuesto por el psicólogo.

b) A partir del nivel de significación α, podemos obtener el nivel crítico y el criterio de decisión :

α = Pr (Rechazar H_0 / H_0 es cierta) = $\mathbf{Pr(\bar{x} > \pi / \mu = 55)} = \Pr\left(z > \dfrac{\pi - 55}{11/\sqrt{50}} \right) = 0'01$

De aquí , consultando las tablas de la distribución normal : $\mathbf{\dfrac{\pi - 55}{11/\sqrt{50}} = 2'33 \quad \Rightarrow \quad \pi = 58'62}$

El criterio de decisión es ***"Aceptamos H_0 si la media muestral es inferior a 58'62"*** , lo cuál concuerda con la conclusión del apartado **a)** ya que al ser la media muestral 58 < 58'62 nos permite aceptar la hipótesis nula.

Conocido el nivel crítico π, estamos en condiciones de calcular β :

β = Pr (Aceptar H_0 / H_0 es falsa) = $\mathbf{Pr(\bar{x} > \pi / \mu = 60)} = \Pr\left(z < \dfrac{58'62 - 60}{11/\sqrt{50}} \right)$

$$\beta = \Pr(z < -0'89) = 0'1867$$

c) Potencia del contraste = η = 1 - β = 1- 0'1867 = 0'8133.
Incrementando η en un 10%, la nueva potencia será : η = 0'8133 + 0'10 . 0'8133 = 0'8946
El nuevo valor de β es : β = 1 - η = 1 - 0'8946 = 0'1054.

Llegado a este punto nos planteamos el modo de cálculo del tamaño muestral solicitado. El enunciado del problema no es lo suficientemente preciso para decidir el criterio de resolución. Puede ser interpretado de dos modos distintos :

1º Determinación del tamaño n, partiendo del valor de β y al margen del valor que correspondería para α :

β = Pr (Aceptar H_0 / H_0 es falsa) = $\mathbf{Pr(\bar{x} > \pi / \mu = 60)} = \Pr\left(z < \dfrac{58'62 - 60}{11/\sqrt{n}} \right) = 0'1054 \Rightarrow$

$$\Rightarrow \quad \mathbf{\frac{58'62 - 60}{11/\sqrt{n}} = -1'25 \quad \Rightarrow \quad \sqrt{n} = \frac{-1'25.11}{58'62 - 60} = 9'9638 \quad \Rightarrow \quad n = 99'28 \cong 99}$$

2º Determinación del tamaño n, partiendo del valor calculado de β,manteniendo α = 0'01 :

$z_\alpha = z_{0'01} = 2'33$
$z_\beta = z_{0'1054} = 1'25$

$$n = \frac{\sigma^2.\left(|z_\alpha| + |z_\beta|\right)^2}{(\mu_0 - \mu_1)^2} = \frac{121.(2'33 + 1'25)^2}{(55 - 60)^2} = 62$$

8

Las determinaciones de colesterol en sangre se distribuyen normalmente con media 180 mg./dl. y varianza 450. De experimentaciones en la zona costera de la provincia de Cádiz se tiene la creencia de que esta cifra media es de 192 mg./dl., por lo que se decide analizar la validez de dicha hipótesis, estableciendo como regla de decisión :

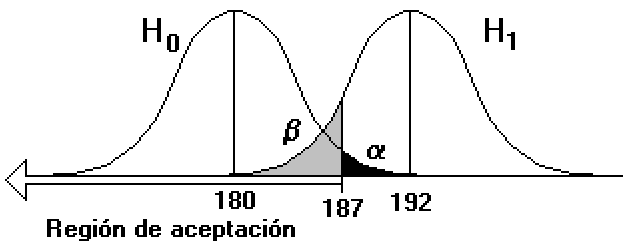
"Se aceptará el promedio 180, establecido como estándar, si la media muestral observada en 80 individuos es inferior a 187 mg./dl.".

a) Calcular las probabilidades de error de tipo I y de tipo II.
b) Trabajando con un nivel de significación del 5%, ¿ qué tamaño muestral se debe tomar ?.

Contraste de la media poblacional. Varianza de la población conocida.

Hipótesis nula y alternativa :

H_0 : μ = 180
H_1 : μ = 192



Función de decisión : La varianza de la población se conoce (σ^2 = 450).

El estadístico de contraste se distribuye normalmente , siendo : $\mathbf{z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}}$

a) α = Pr (Rechazar H_0 / H_0 es cierta) = $\mathbf{Pr(\bar{x} > 187 / \mu = 180)}$

$$\alpha = \Pr\left(z > \frac{187 - 180}{\sqrt{450}/\sqrt{80}} \right) = \Pr(z > 2'95) = 0'00159$$

β = Pr (Aceptar H_0 / H_0 es falsa) = $\mathbf{Pr(\bar{x} < 187 / \mu = 192)}$

$$\beta = \Pr\left(z < \frac{187 - 192}{\sqrt{450}/\sqrt{50}} \right) = \Pr(z < -2'11) = 0'01743$$

b) Determinaremos el tamaño muestral **n** sabiendo que $\boldsymbol{\alpha}$ **= 0'05** , pero ¿ qué valor de $\boldsymbol{\beta}$ utilizamos ?.

1º **Para al valor obtenido en el apartado anterior : β = 0'01743 :**

$z_\alpha = z_{0'05} = 1'64$
$z_\beta = z_{0'01743} = 2'11$

$$\mathbf{n} = \frac{\sigma^2 . \left[|z_\alpha| + |z_\beta| \right]^2}{(\mu_0 - \mu_1)^2} = \frac{450.(1'64 + 2'11)^2}{(180 - 192)^2} = 43'95 \cong 44$$

2º **Al margen del valor de β que correspondería para α = 0'05 , y manteniendo el criterio de decisión :**

α = Pr (Rechazar H_0 / H_0 es cierta) = $\mathbf{Pr(\bar{x} > 187 / \mu = 180)}$ $= \Pr\left(z > \frac{187 - 180}{\sqrt{450}/\sqrt{n}} \right) = 0'05$

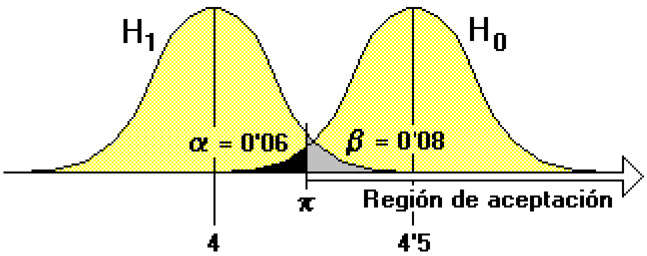
$$\frac{187 - 180}{\sqrt{450}/\sqrt{n}} = 1'64 \quad \Rightarrow \quad \sqrt{n} = \frac{1'64.\sqrt{450}}{187 - 180} = 4'97 \quad \Rightarrow \quad n = 24'7 \cong 25$$

9

Las calificaciones de la pruebas de acceso a la Universidad para mayores de 25 años se ajustan a una distribución normal N(4'5 , 0'9). El análisis de los resultados obtenidos en las pasadas convocatorias hacen pensar que el promedio adecuado es 4.
a) ¿ Qué tamaño muestral debe utilizarse para que las probabilidades de los errores de tipo I y II sean iguales a 0'06 y 0'08 respectivamente ?.
b) Si, seleccionada una muestra de dicho tamaño, se obtiene un promedio 4'2, ¿ qué decisión debemos adoptar ?.

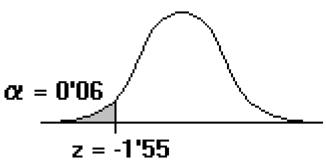
Contraste de la media poblacional. Varianza de la población conocida ($\sigma^2 = 0'9^2 = 0'81$).

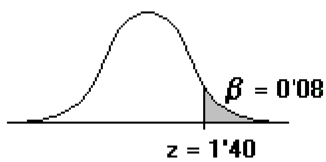
$H_0 : \ \mu = 4'5$
$H_1 : \ \mu = 4$



Función de decisión : $\mathbf{z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}}$ se distribuye normalmente.

a) Cálculo del tamaño muestral **n** :





$z_\alpha = z_{0'06} = -1'55$
$z_\beta = z_{0'08} = 1'40$

$$n = \frac{\sigma^2 . \left(\left|z_\alpha\right| + \left|z_\beta\right|\right)^2}{(\mu_0 - \mu_1)^2} = \frac{0'9^2.(1'55 + 1'40)^2}{(4'5 - 4)^2} = 28'97 \cong 29$$

b) Para n = 29, el estadístico de contraste toma el valor :

$$\mathbf{z} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{4'2 - 4'5}{0'9 / \sqrt{29}} = -1'80$$

Con el nivel de significación $\alpha = 0'06$, la tabla de la distribución N(0,1) proporcionó el nivel crítico de decisión :

$z_\alpha = z_{0'06} = -1'55$

Siendo $z < z_\alpha$ (-1'80 < -155) rechazamos la hipótesis nula, aceptando que la media más adecuada es 4.

10

Los ingresos mensuales por dietas de desplazamiento de los trabajadores de una empresa, se calculó que tenían como media 32000 pts. y desviación típica 8000.
Si se establece en 35000 el nivel crítico de decisión y seleccionamos muestras de 16 individuos, calcule las potencias de los contrastes con alternativas sucesivas :

37000 38000 39000 40000

Contraste de la media poblacional. Varianza de la población conocida.

Función de decisión : $\mathbf{z} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ distribuida según una N(0,1).

$H_0 : \mu = 32000$
$H_1 : \mu = 37000$

β = Pr (Aceptar H_0 / H_0 es falsa) = $\mathbf{Pr(\bar{x} < 35000 / \mu = 37000)}$

$$\beta = \Pr\left(z < \frac{35000 - 37000}{8000/\sqrt{16}}\right) = \Pr(z < -1) = 0'15866$$

η = 1 - β = 1 - 0'15866 = 0'84134

$H_0 : \mu = 32000$
$H_1 : \mu = 38000$

β = Pr (Aceptar H_0 / H_0 es falsa) = $\mathbf{Pr(\bar{x} < 35000 / \mu = 38000)}$

$$\beta = \Pr\left(z < \frac{35000 - 38000}{8000/\sqrt{16}}\right) = \Pr(z < -1'5) = 0'06681$$

η = 1 - β = 1 - 0'06681 = 0'93319

$H_0 : \mu = 32000$
$H_1 : \mu = 39000$

β = Pr (Aceptar H_0 / H_0 es falsa) = $\mathbf{Pr(\bar{x} < 35000 / \mu = 39000)}$

$$\beta = \Pr\left(z < \frac{35000 - 39000}{8000/\sqrt{16}}\right) = \Pr(z < -2) = 0'02275$$

η = 1 - β = 1 - 0'02275 = 0'97725

$H_0 : \mu = 32000$
$H_1 : \mu = 40000$

β = Pr (Aceptar H_0 / H_0 es falsa) = $\mathbf{Pr(\bar{x} < 35000 / \mu = 40000)}$

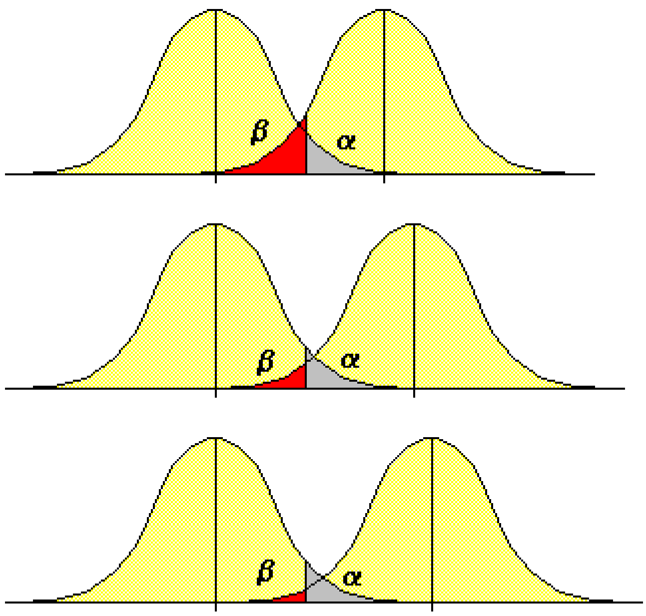
$$\beta = \Pr\left(z < \frac{35000 - 40000}{8000/\sqrt{16}}\right) = \Pr(z < -2'5) = 0'00621$$

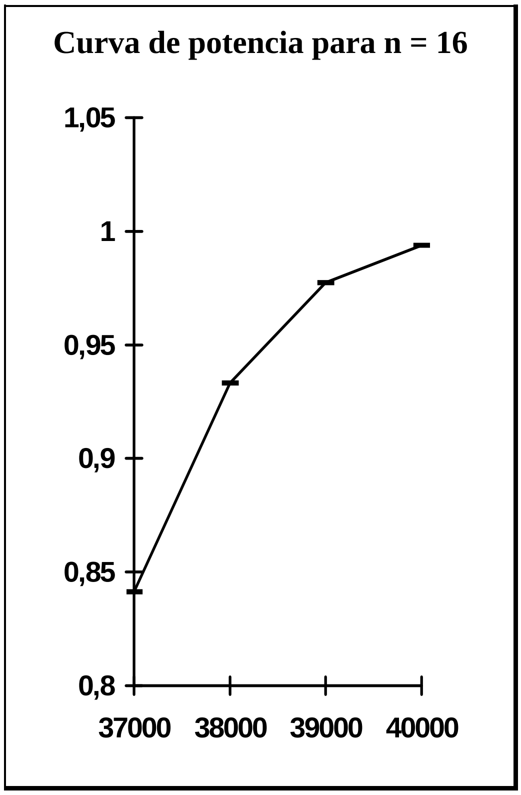
η = 1 - β = 1 - 0'00621 = 0'99379

Habiendo fijado el criterio de decisión del contraste, lo cuál es equivalente a establecer el nivel de significación α o probabilidad del error de tipo I, al alejar la hipótesis alternativa hacemos que la probabilidad del error de tipo II disminuya.

Al diminuir β, la potencia del contraste (η = 1-β) aumentará. La gráfica de valores sucesivos de η recibe el nombre de "***curva de potencia***". La construimos a continuación.

Representación gráfica de la conclusión anterior :





11

Los ingresos mensuales por dietas de desplazamiento de los trabajadores de una empresa, se calculó que tenían como media 32000 pts. y desviación típica 8000.
Si se establece en 35000 el nivel crítico de decisión para contrastar la hipótesis con alternativa 38000, calcule las potencias correspondientes a la selección de muestras de tamaños :

10 40 70 100

Contraste de la media poblacional. Varianza de la población conocida.

Función de decisión : $\mathbf{z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}}$ distribuida según una N(0,1).

$H_0: \ \mu = 32000$
$H_1: \ \mu = 38000$

n = 10

β = Pr (Aceptar H_0 / H_0 es falsa) = $\mathbf{Pr(\bar{x} < 35000 / \mu = 38000)}$

$$\beta = \Pr\left(z < \frac{35000 - 38000}{8000/\sqrt{10}} \right) = \Pr(z < -1'19) = 0'11702$$

η = 1 - β = 1 - 0'11702 = 0'88298

$H_0: \ \mu = 32000$
$H_1: \ \mu = 38000$

n = 40

β = Pr (Aceptar H_0 / H_0 es falsa) = $\mathbf{Pr(\bar{x} < 35000 / \mu = 38000)}$

$$\beta = \Pr\left(z < \frac{35000 - 38000}{8000/\sqrt{40}} \right) = \Pr(z < -2'37) = 0'00889$$

η = 1 - β = 1 - 0'00889 = 0'99111

$H_0: \ \mu = 32000$
$H_1: \ \mu = 38000$

n = 70

β = Pr (Aceptar H_0 / H_0 es falsa) = $\mathbf{Pr(\bar{x} < 35000 / \mu = 38000)}$

$$\beta = \Pr\left(z < \frac{35000 - 38000}{8000/\sqrt{70}} \right) = \Pr(z < -3'14) = 0'00085$$

η = 1 - β = 1 - 0'00085 = 0'99915

$H_0: \ \mu = 32000$
$H_1: \ \mu = 38000$

$n = 100$

β = Pr (Aceptar H_0 / H_0 es falsa) = $\mathbf{Pr(\bar{x} < 35000 \,/\, \mu = 38000)}$

$$\beta = \Pr\left(z < \frac{35000-38000}{8000/\sqrt{100}} \right) = \Pr(z < -3'75) = 0'00009$$

$$\boldsymbol{\eta = 1 - \beta = 1 - 0'00009 = 0'99991}$$

Fijando el criterio de decisión del contraste, al aumentar el tamaño muestral conseguimos que la probabilidad del error de tipo II disminuya y, en consecuencia, que el contraste sea más potente.

12

En una muestra aleatoria de 16 alumnos de 2º de E.G.B. de un determinado Colegio, hemos obtenido una media igual a 105 en un test de inteligencia. La suma de cuadrados de las puntuaciones diferenciales respecto a la media resultó ser igual a 240.
A continuación, se toma una muestra de 14 alumnas del mismo curso y Centro escolar y se obtiene una media igual a 102 en el mismo test, y una desviación típica insesgada de 3.
¿ Podemos afirmar con un nivel de riesgo del 5% que los alumnos y alumnas de 2º de E.G.B. de ese Centro poseen el mismo nivel de inteligencia ?.

Contraste relativo a una igualdad de medias poblacionales, en muestras independientes.

Hipótesis :

$H_0: \ \mu_1 = \mu_2 \qquad (\mu_1 - \mu_2 = 0)$
$H_1: \ \mu_1 \neq \mu_2 \qquad (\mu_1 - \mu_2 \neq 0)$

Al no realizar el enunciado del problema indicación alguna sobre las varianzas de la población, consideramos el caso más amplio : varianzas poblacionales desconocidas y distintas.

Calculemos previamente el valor de los grados de libertad, para lo cuál determinaremos antes las varianzas muestrales.

$$n_1 = 16 \quad ; \quad \sum (x-\bar{x})^2 = 240 \quad \Rightarrow \quad s_1^2 = \frac{\sum (x-\bar{x})^2}{n_1} = \frac{240}{16} = 15$$

$$n_2 = 14 \quad ; \quad s'_2 = 3 \quad \Rightarrow \quad s_2^2 = \frac{n_1 - 1}{n_1} . s'^2_2 = \frac{13}{14} . 9 = 8'357$$

$$gl = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right]^2}{\frac{\left(s_1^2/n_1 \right)^2}{n_1+1} + \frac{\left(s_2^2/n_2 \right)^2}{n_2+1}} - 2 = \frac{\left[\frac{15}{16} + \frac{8'357}{14} \right]^2}{\frac{\left(15/16\right)^2}{17} + \frac{\left(8'357/14\right)^2}{15}} - 2 = 29'204 \approx 29$$

Al ser **gl = 29 ≤ 50**, la **función de decisión** es :

$$t_{gl,\alpha} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}} = \frac{(105-102)}{\sqrt{\frac{15}{15} + \frac{8'357}{13}}} = 2'3406$$

Para un nivel de significación del 5%, la tabla t de Student proporciona los valores : $t_{29,0'05/2} = \pm 2'045230$.

Dado que el valor calculado 2'3406 queda fuera del intervalo (-2'045230 , 2'045230) , rechazamos la hipótesis nula. En consecuencia, no podemos admitir que los alumnos y las alumnas del Centro posean el mismo nivel de inteligencia.

13

Un centro que imparte clases de E.G.B. tiene distribuidos a los 55 alumnos de 5º en dos grupos, A y B, de 25 y 30 alumnos respectivamente. Al acabar el curso aprueban 17 alumnos del grupo A y 18 del B.
Con un nivel de confianza del 95%, ¿ podemos afirmar estadísticamente que superan el curso el mismo número de alumnos en ambos grupos ?, o bien, ¿ es independiente el hecho de superar el curso con respecto al grupo en el que esté ?.

Contraste relativo a una igualdad de proporciones poblacionales, en muestras independientes.

Las proporciones muestrales observadas son :

A) $n_1 = 25 \quad ; \quad p_1 = \frac{17}{25} = 0'68$ B) $n_2 = 30 \quad ; \quad p_2 = \frac{18}{30} = 0'6$

Hipótesis :

$$H_0: \ \pi_1 = \pi_2 \qquad (\pi_1 - \pi_2 = 0)$$
$$H_1: \ \pi_1 \neq \pi_2 \qquad (\pi_1 - \pi_2 \neq 0)$$

Siendo :

$$D = \sqrt{\left(\frac{n_1.p_1 + n_2.p_2}{n_1 + n_2}\right).\left(1 - \frac{n_1.p_1 + n_2.p_2}{n_1 + n_2}\right).\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$
$$= \sqrt{\left(\frac{25.0'68 + 30.0'6}{25 + 30}\right).\left(1 - \frac{25.0'68 + 30.0'6}{25 + 30}\right).\left(\frac{1}{25} + \frac{1}{30}\right)} = 0'1303$$

la **función de decisión** es :

$$z_\alpha = \frac{p_1 - p_2}{D} = \frac{0'68 - 0'6}{0'1303} = 0'614$$

La tabla normal proporciona, para un nivel de significación del 5% (confianza del 95%) : $z_\alpha = \pm 1'96$.

Dado que el valor calculado 0’614 pertenece al intervalo (-1’96 , 1’96) , aceptamos la hipótesis nula. En consecuencia, el curso es superado por el mismo número de alumnos en ambos grupos.

14

Un centro escolar tiene matriculados en C.O.U. 250 alumnos, de los que 110 fuman habitual u ocasionalmente. Tras realizar un cursillo sobre el tema, han decidido dejar de fumar el 60% de los que lo hacían, pero un 10% de los que no fumaban les ha picado la curiosidad y han comenzado a fumar.
¿ Podemos afirmar, con un nivel de confianza del 95%, que el cursillo ha conseguido el fin que perseguía ?.

Hipótesis :

$$H_0: \ \pi_1 = \pi_2 \qquad (\pi_1 - \pi_2 = 0)$$
$$H_1: \ \pi_1 \neq \pi_2 \qquad (\pi_1 - \pi_2 \neq 0)$$

<u>Muestras relacionadas</u>

Disponemos las frecuencias observadas antes y después del cursillo en la tabla siguiente :

		Después		
		Fuma	**No fuma**	
Antes	**No fuma**	**A=14**	**B=126**	**140**
	Fuma	**C=44**	**D=66**	**110**
		58	**192**	**250**

Función de decisión :

$$z_\alpha = \frac{D - A}{\sqrt{A + D}} = \frac{66 - 14}{\sqrt{14 + 66}} = 5'814$$

La tabla normal proporciona, para un nivel de significación del 5% (confianza del 95%) : $z_\alpha = \pm 1'96$.

Dado que el valor calculado 5’814 no pertenece al intervalo (-1’96 , 1’96) , rechazamos la hipótesis nula. En consecuencia, existe influencia significativa del cursillo en el cambio de hábito de fumar.

15

Diferentes estudios realizados en el campo del procesamiento de la información, muestran que las personas retrasadas la procesan más lentamente que las personas normales. En esta línea, queremos averiguar si las personas retrasadas presentan un tiempo de inspección (una medida de la rapidez del procesamiento sensorial) más largo que las personas normales, lo que indicaría un procesamiento sensorial más deficiente.
Para ello elegimos una muestra aleatoria de 8 sujetos retrasados y otra de 8 normales. Los tiempos de inspección (TI) de las dos muestras de sujetos son los siguientes :

Retrasados	**150**	**180**	**300**	**240**	**170**	**220**	**280**	**195**
Normales	**110**	**120**	**90**	**125**	**122**	**80**	**115**	**95**

Suponiendo que las dos distribuciones se distribuyen normalmente y las dos varianzas poblacionales son distintas, aplique una técnica paramétrica para someter a prueba la hipótesis de que el TI de las personas retrasadas es más largo que el de las normales.

Contraste relativo a una comparación de medias poblacionales, en muestras independientes.

Hipótesis :

$$H_0: \ \mu_1 > \mu_2 \qquad (\mu_1 - \mu_2 > 0) \qquad \text{(retrasados mayor que normales)}$$
$$H_1: \ \mu_1 \leq \mu_2 \qquad (\mu_1 - \mu_2 \leq 0)$$

Parámetros muestrales :

(1) Retrasados	150	180	300	240	170	220	280	195	Σ = 1735
cuadrados	22500	32400	90000	57600	28900	48400	78400	38025	Σ = 396225

(2) Normales	110	120	90	125	122	80	115	95	Σ = 857
cuadrados	12100	14400	8100	15625	14884	6400	13225	9025	Σ = 93759

$$\bar{x}_1 = \frac{1735}{8} = 216'875 \qquad s_1^2 = \frac{396225}{8} - 216'875^2 = 2493'36$$

$$\bar{x}_2 = \frac{857}{8} = 107'125 \qquad s_2^2 = \frac{93759}{8} - 107'125^2 = 244'11$$

Función de decisión :

Se trata de varianzas poblacionales desconocidas y distintas, con muestras pequeñas del mismo tamaño (n = 8).

$$t_{2n-2,\alpha} = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{s_1^2 + s_2^2}{n-1}}} = \frac{216'875 - 107'125}{\sqrt{\dfrac{2493'36 + 244'11}{8-1}}} = 5'55$$

Nos encontramos ante un contraste unilateral. La hipótesis alternativa (μ_1 - $\mu_2 \leq 0$) establece que el nivel de significación (α) quedará en su totalidad a la izquierda.
El intervalo o región de aceptación será del tipo : **($-t_{2n-2,\alpha}$, $+\infty$)**

Fijando α = 0'05 y, para 2n-2 = 14 grados de libertad, obtenemos de la tabla t de Student (unilateral) : $t_{14,0'05}$ = 1'761304

Como **5'55** pertenece al intervalo **(-1'761304 , +∞)** , aceptamos la hipótesis nula. Luego, el TI de las personas retrasadas es más largo que el de las normales.

16

Los siguientes datos han sido tomados de una investigación de Campbell, representando las puntuaciones de 10 sujetos en una escala de impulsividad :

Sujeto	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Impulsividad	**2**	**3**	**4**	**16**	**9**	**2**	**6**	**12**	**2**	**8**

Utilizando un nivel de significación igual a 0'05, aplique una técnica paramétrica para contrastar la hipótesis nula de que la varianza poblacional de las puntuaciones en la escala de impulsividad es igual a 36.

Contraste de la varianza poblacional.

Hipótesis :

H_0 : $\sigma^2 = 36$
H_1 : $\sigma^2 \neq 36$ (Contraste bilateral)

Parámetros muestrales :

Impulsividad (x)	2	3	4	16	9	2	6	12	2	8	Σ = 64
x^2	4	9	16	256	81	4	36	144	4	64	Σ = 618

$$\bar{x} = \frac{64}{10} = 6'4 \qquad s^2 = \frac{618}{10} - 6'4^2 = 20'84$$

Método 1º, basado en la función de decisión :

$$\chi^2_{n-1,\alpha} = \frac{n.s^2}{\sigma^2} = \frac{10\,.\,20'84}{36} = 5'789$$

Consultando la tabla de χ^2 con n-1 = 9 grados de libertad, obtenemos los valores que dejan a derecha y a izquierda un área $\alpha/2$ = 0'025 . Estos son : 2'700439 y 19'022827

El intervalo o región de aceptación es pues : **(2'700439 , 19'022827)** .

Como el valor obtenido en la función de decisión (**5'789**) se encuentra dentro del intervalo **(2'700439 , 19'022827)**, aceptamos la hipótesis de que la varianza poblacional es igual a 36.

Método 2º, basado en la región de aceptación :

$$\left(\frac{\sigma^2.\chi_1^2}{n},\frac{\sigma^2.\chi_2^2}{n}\right)=\left(\frac{36.2'700439}{10},\frac{36.19'022827}{10}\right)=(9'72\ ,\ 68'4828)$$

Como la varianza muestral (**20’84**) pertenece a la región de aceptación de la hipótesis nula **(9’72 , 68’4828)** , mantenemos que la varianza poblacional es igual a 36.

17

En octubre de 1991 se celebraron unas vistas en el Senado de Estados Unidos sobre un caso de acoso sexual, presentado por Anita Hill (AH) contra el juez Clarence Thomas (CT) que afectaría, o no, su nominación como juez del Tribunal Supremo. El juez fue confirmado para el Supremo. El periódico “Los Ángeles Times” ha realizado una encuesta para analizar qué grupos poblacionales de mujeres creyeron el testimonio de AH. Los resultados obtenidos fueron los siguientes :

- **El 32% de las mujeres con estudios superiores (ES) creyeron el testimonio de AH versus un 18% de las que no tenían estudios superiores (SES).**
- **El 54% de las mujeres que trabajan fuera de casa (T) creyeron el testimonio de AH versus un 44% de las que no trabajan (NT).**
- **El 60% de las mujeres con categorías profesionales altas (CA) creyeron el testimonio de AH versus un 45% de las que tenían categorías profesionales bajas (CB).**

Asumiendo que las mujeres se seleccionaron al azar, que las observaciones y las muestras son independientes, que el tamaño muestral fue n = 100 en todos los casos y, que el margen de error es del 5%, ¿ entre qué pares posibles existen diferencias significativas ?.

Se trata de tres contrastes sobre igualdad de proporciones.

Hipótesis :

$H_0:\ \pi_1 = \pi_2$ $\qquad(\pi_1 - \pi_2 = 0)$ $\qquad$ **<u>Muestras independientes</u>**

$H_1:\ \pi_1 \neq \pi_2$ $\qquad(\pi_1 - \pi_2 \neq 0)$

Función de decisión :

$$z_\alpha = \frac{p_1 - p_2}{D} \quad \text{siendo} \quad D=\sqrt{\left(\frac{n_1.p_1+n_2.p_2}{n_1+n_2}\right).\left(1-\frac{n_1.p_1+n_2.p_2}{n_1+n_2}\right).\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Región de aceptación :

Para un nivel de significación α = 0’05, la tabla normal proporciona los valores $z_\alpha = \pm1'96$. Es decir, aceptaremos que no existen diferencias significativas si el valor de la función de decisión está en el intervalo **(-1’96 , 1’96)**.

A) Grupo en función de haber cursado o no estudios superiores :

(ES) $p_1 = 0'32$; **(SES)** $p_2 = 0'18$; $n_1 = n_2 = 100$; $n_1.p_1 = 32$; $n_2.p_2 = 18$

$$D=\sqrt{\left(\frac{32+18}{100+100}\right).\left(1-\frac{32+18}{100+100}\right).\left(\frac{1}{100}+\frac{1}{100}\right)}=0'06124$$

$$z_\alpha = \frac{0'32-0'18}{0'06124}=2'286 \quad \notin \quad (-1'96, 1'96)$$. Existen diferencias significativas.

B) Grupo en función de trabajar fuera de casa :

(T) $p_1 = 0'54$; **(NT)** $p_2 = 0'44$; $n_1 = n_2 = 100$; $n_1.p_1 = 54$; $n_2.p_2 = 44$

$$D=\sqrt{\left(\frac{54+44}{100+100}\right).\left(1-\frac{54+44}{100+100}\right).\left(\frac{1}{100}+\frac{1}{100}\right)}=0'07069$$

$$z_\alpha = \frac{0'54-0'44}{0'07069}=1'415 \quad \in \quad (-1'96, 1'96)$$. No existen diferencias significativas.

C) Grupo en función de la categoría profesional :

(CA) $p_1 = 0'60$; **(CB)** $p_2 = 0'45$; $n_1 = n_2 = 100$; $n_1.p_1 = 60$; $n_2.p_2 = 45$

$$D=\sqrt{\left(\frac{60+45}{100+100}\right).\left(1-\frac{60+45}{100+100}\right).\left(\frac{1}{100}+\frac{1}{100}\right)}=0'07062$$

$$z_\alpha = \frac{0'60-0'45}{0'07062}=2'124 \quad \notin \quad (-1'96, 1'96)$$. Existen diferencias significativas.

18

Una empresa de transporte de mercancías tiene dos oficinas (A y B) y desean analizar el tiempo de llamadas por FAX que recibe cada una de ellas con el fin de distribuir los aparatos de acuerdo con el tiempo de utilización de los mismos. Para ello, seleccionaron al azar 7 líneas de FAX en cada oficina y registraron los tiempos de utilización en minutos a lo largo de un día en la tabla siguiente :

Oficina A	**120**	**150**	**110**	**200**	**140**	**230**	**170**
Oficina B	**130**	**160**	**120**	**210**	**170**	**250**	**190**

Utilizando un nivel de significación 0’01 y suponiendo que las distribuciones son normales :

a) Aplique una técnica paramétrica para someter a prueba la hipótesis de que las varianzas de los tiempos de utilización del FAX en las dos oficinas son iguales.

b) Basándose en el resultado del apartado anterior, aplique otra técnica paramétrica para someter a prueba la hipótesis de que el tiempo medio de utilización del FAX es menor en la oficina A que en la B.

a) Comparación de dos varianzas para muestras independientes.

Hipótesis : (Contraste bilateral)

$$H_0:\quad \sigma_1^2 = \sigma_2^2 \qquad \left(\frac{\sigma_1^2}{\sigma_2^2} = 1\right) \qquad\qquad H_1:\quad \sigma_1^2 \neq \sigma_2^2 \qquad \left(\frac{\sigma_1^2}{\sigma_2^2} \neq 1\right)$$

Parámetros muestrales :

Calculamos las cuasivarianzas a partir de las siete observaciones de la oficina A y la B. Estos cálculos conducen a :

$$\overline{x}_A = \overline{x}_1 = 160 \qquad s_A^2 = s_1^2 = 1600'86 \qquad s'^2_A = s'^2_1 = 1867'67$$

$$\overline{x}_B = \overline{x}_2 = 175'71 \qquad s_B^2 = s_2^2 = 1767'34 \qquad s'^2_B = s'^2_2 = 2061'90$$

Función de decisión :

$$F_{n_1-1,n_2-1,\alpha} = \frac{s_1'^2}{s_2'^2}\cdot\frac{\sigma_2^2}{\sigma_1^2} = \frac{1867'67}{2061'90}.1 = 0'9058$$

Correspondiendo a una F de Snedecor con 6 y 6 (n_1-1 y n_2-1) grados de libertad, la tabla relativa a

$\alpha/2 = 0'01/2 =$ **0’005** (contraste bilateral)

proporciona los valores tabulados (límites de la región de aceptación) :

$$F_{6,6,0'005} = 11'070313 \qquad F_{6,6,1-0'005} = F_{6,6,0'995} = \frac{1}{F_{6,6,0'005}} = \frac{1}{11'070313} = 0'090332$$

Al pertenecer 0’9058 al intervalo (0’090332 , 11’070313), aceptamos la hipótesis de igualdad de dispersiones (varianzas).

b) Comparación de dos medias para muestras independientes.

Hipótesis : (Contraste unilateral)

$H_0:\ \mu_1 < \mu_2$ $\quad(\mu_1 - \mu_2 < 0)$ $\quad$ (oficina A menor que la B)

$H_1:\ \mu_1 \geq \mu_2$ $\quad(\mu_1 - \mu_2 \geq 0)$

Función de decisión :

Al ser las varianzas poblacionales desconocidas pero iguales (según el apartado anterior), las muestras independientes y de tamaño pequeño, la función de decisión es :

$$t_{n_1+n_2-2,\alpha} = \frac{\left(\overline{x}_1 - \overline{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{n_1 . s_1^2 + n_2 . s_2^2}{n_1 + n_2 - 2}}.\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{160 - 175'71}{\sqrt{\frac{7.1600'86 + 7.1767'34}{7+7-2}}.\sqrt{\frac{1}{7} + \frac{1}{7}}} = -0'663$$

Al tratarse de un contraste unilateral la significación α=0’01 quedará en su totalidad a la derecha (alternativa : $\mu_1 - \mu_2 \geq 0$).

La tabla t de Student, para 12 grados de libertad (7+7-2), proporciona : $t_{12,0'01} = 2'680955$.

Puesto que el valor **-0’663** de la función de decisión se encuentra en el intervalo **(-∞ , 2’680955)** , de aceptación de la hipótesis nula, concluimos que el tiempo medio de utilización del FAX es menor en la oficina A que en la B.

19

Garriga-Trillo, González-Labra, Villarino, Lubín, García-Gallego & Arnau (1994) y Baird & Harder (1994) realizaron un estudio comparativo entre una muestra española (E) y otra estdounidense (EU), considerando variables de tipo psicofísico (entre ellas la sensibilidad visual). En ambos casos se utilizó una muestra aleatoria simple de 50 sujetos. Al analizar los datos se encontró que la variabilidad, medida por la desviación típica insesgada, de las puntuaciones en sensibilidad visual (S) en E fue de 0'2 y en EU de 0'3. Asumiendo la normalidad de las puntuaciones en sensibilidad visual en ambas poblaciones, se quiere contrastar si la variabilidad en EU es realmente superior a la de E. Para ello, conteste :

a) ¿ Cuál es la hipótesis nula y cuál la alternativa ?.
b) Elija el estadístico a utilizar, justificando su elección, para contrastar si existen diferencias significativas en la variabilidad de las puntuaciones en sensibilidad visual.
c) Calcúlelo y determine la probabilidad aproximada (p) de obtener ese valor o uno menor.
d) Utilizando ese valor p como nivel de significación, interprete el resultado obtenido en el contexto de la investigación.

a)

$$H_0: \quad \sigma_E^2 < \sigma_{EU}^2 \qquad \left(\frac{\sigma_E^2}{\sigma_{EU}^2} < 1\right) \qquad\qquad H_1: \quad \sigma_E^2 \geq \sigma_{EU}^2 \qquad \left(\frac{\sigma_E^2}{\sigma_{EU}^2} \geq 1\right)$$

b)

Nos encontramos ante un contraste relativo a la comparación de dos varianzas en dos muestras independientes. El estadístico de contraste es :

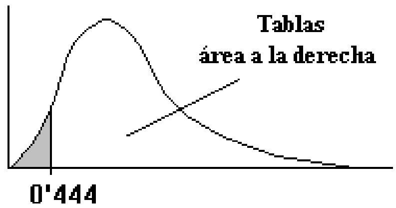
$$F_{n_1-1,n_2-1,\alpha} = \frac{s_1'^2}{s_2'^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \qquad (E = 1 \,;\, EU = 2)$$

c)

Sustituyendo los valores muestrales en la función de decisión anterior, obtenemos :

$$F = \frac{0'2^2}{0'3^2} = 0'444$$

Calculemos la probabilidad : $p = Prob(\, F_{49,49} \leq 0'444)$:



Hemos de recorrer las distintas tablas de la F de Snedecor obteniendo los valores F correspondientes a 49, 49 grados de libertad, así como sus inversos (áreas complementarias 1-α) :

área = α = 0'05	F = 1'607300
área = α = 0'025	F = 1'762207
área = α = 0'01	F = 1'962646
área = α = 0'005	F = 2'113037
área = 1-α = 0'95	F = 1/1'607300 = 0'622161
área = 1-α = 0'975	F = 1/1'762207 = 0'567470
área = 1-α = 0'99	F = 1/1'962646 = 0'509516
área = 1-α = 0'995	F = 1/2'113037 = 0'473252

El valor más próximo a 0'444 lo encontramos en 0'473252, el cuál deja a su derecha un área 0'995, luego dejará a su izquierda un área 0'005 que es la probabilidad pedida :

$$\mathbf{p = Prob(\, F_{49,49} \leq 0'444) \approx 0'005}$$

d)

Con un nivel de significación $\alpha = p = 0'005$ y siendo el contraste unilateral , la hipótesis alternativa nos indica que el área α quedará en su totalidad a la derecha.

La región de aceptación es pues : $(-\infty \,,\, F_{49,49,0'005}) = (-\infty \,,\, 2'113037)$

Como $0'444 \in (-\infty \,,\, 2'30)$, aceptamos la hipótesis nula de que la variabilidad en EU es superior a la de E.

NOTA : $\frac{\sigma_{EU}^2}{\sigma_E^2} > 1$	La hipótesis nula pudo haber sido planteada en la forma que indica la expresión de la izquierda. En este caso, el estadístico F de contraste habría tomado el valor : $F = \frac{0'3^2}{0'2^2} = 2'25$ Con ello, el cálculo de la probabilidad p, habría considerado el valor tabulado 2'113037 como más próximo (área a la derecha α = 0'005), dando como resultado : **p = Prob($F_{49,49} \leq$ 2'25) ≈ 0'995** (1 - α = 1 - 0'005) Para el apartado **d)**, resultaría inapropiado tomar un nivel de significación $\alpha = p = 0'995$. La lógica nos debe llevar a tomar su complemento a la unidad; es decir : $\alpha = 1 - p = 0'005$

20

Se desea estudiar la eficacia de determinada terapia para el control de la agresividad. Para ello, se seleccionó una muestra aleatoria de 7 sujetos violentos y se les aplicó una escala de agresividad (pretest). Posteriormente se les trató durante unos meses con dicha terapia y, a continuación, se les volvió a pasar la escala de agresividad (postest). Aquí aparecen las medidas pretest y postest obtenidas :

Pretest	**30**	**34**	**50**	**40**	**25**	**28**	**46**
Postest	**10**	**25**	**36**	**22**	**20**	**19**	**30**

a) Sabiendo que las puntuaciones en la escala de agresividad son una variable continua y que las n diferencias (medida pretest menos medida postest) son independientes y se distribuyen normalmente, ¿ cuál es el contraste más apropiado para comprobar la eficacia de la terapia ?. Razone su elección, formule las hipótesis estadísticas, aplique dicho contraste utilizando un α = 0’01, e interprete el resultado.

b) Sabiendo que las varianzas sesgadas de las medidas pretest y postest fueron, respectivamente, 76’69 y 59’55, y que la correlación entre las dos series de puntuaciones es $r_{pre\text{-}pos}$ = 0’82, contraste si existe homogeneidad entre los resultados obtenidos en las dos medidas. Formule las hipótesis, aplique el estadístico más adecuado utilizando un α = 0’05, e interprete el resultado.

a) Comparación de medias (varianza de la diferencia desconocida)

Hipótesis :

$H_0: \ \mu_1 > \mu_2 \qquad (\mu_1 - \mu_2 > 0)$ (Eficacia equivale a : Pretest > Postest)

$H_1: \ \mu_1 \le \mu_2 \qquad (\mu_1 - \mu_2 \le 0)$

Parámetro muestrales (de la diferencia) :

Pretest	30	34	50	40	25	28	46
Postest	10	25	36	22	20	19	30

Diferencia (d)	20	9	14	18	5	9	16	Σ = 91
d^2	400	81	196	324	21	81	256	Σ = 1359

$$\overline{D} = \frac{91}{7} = 13 \qquad s_D = \sqrt{\frac{1359}{7} - 13^2} = 5'014 \qquad s'_D = \sqrt{\frac{7}{7-1}}.5'014 = 5'416$$

Función de decisión :

Puede calcularse a partir de la varianza o de la cuasivarianza muestral, siendo el estadístico de contraste :

$$t_{n-1,\alpha} = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{s_D / \sqrt{n-1}} = \frac{13}{5'014 / \sqrt{7-1}} = 6'351 \qquad (\overline{x}_1 - \overline{x}_2 = \overline{D})$$

$$t_{n-1,\alpha} = \frac{\overline{D} - (\mu_1 - \mu_2)}{s'_D / \sqrt{n}} = \frac{13}{5'416 / \sqrt{7}} = 6'351$$

Consultando las tablas de la t de Student (contraste unilateral) obtenemos :

$$\mathbf{t_{7-1,0'05} = 1'942911}$$

El contraste unilateral, con alternativa $\mu_1 - \mu_2 \le 0$, dejará el nivel de significación a la izquierda. El intervalo o región de aceptación será : **(-1’942911 , +∞)**.

Aceptamos la hipótesis nula, de que las puntuaciones pretest son superiores a las postest (eficacia), al pertenecer 6’351 a la región de aceptación **(-1’942911 , +∞)**.

b) Comparación de varianzas (muestras relacionadas).

El término “ser igualmente homogéneas” representa tener la misma dispersión.

Hipótesis :

$$H_0: \ \sigma_1^2 = \sigma_2^2 \quad \left(\frac{\sigma_1^2}{\sigma_2^2} = 1\right) \qquad H_1: \ \sigma_1^2 \ne \sigma_2^2 \quad \left(\frac{\sigma_1^2}{\sigma_2^2} \ne 1\right)$$

Función de decisión :

Necesitamos calcular las varianzas muestrales de las dos series (pretest y postest). Los resultados son :

$$s^2_{pretest} = s_1^2 = 76'69 \qquad s^2_{postest} = s_2^2 = 59'55$$

Con esto :

$$t_{n-2,\alpha} = \frac{(s_1^2 - s_2^2).\sqrt{n-2}}{2.s_1.s_2.\sqrt{1-r^2}} = \frac{(76'69 - 59'55).\sqrt{7-2}}{2.76'69.59'55.\sqrt{1-0'82^2}} = 0'495$$

Consultando las tablas de la t de Student (contraste bilateral) obtenemos :

$$\mathbf{t_{7\text{-}2,0'05} = \pm 2'570796}$$

Aceptamos la hipótesis de igualdad de varianzas de las puntuaciones pretest y postest, al pertenecer 0'495 a la región de aceptación : **(-2'570796 , 2'570796)**.

21

En un estudio sobre los efectos de la contaminación de una sustancia química en malformaciones congénitas, se encontró que el 20% de una muestra aleatoria simple de 90 niños nacidos de madres expuestas accidentalmente a esa sustancia nacieron sin la mano izquierda, mientras que en una muestra aleatoria simple de 90 niños nacidos de madres no expuestas, el 5% nacieron sin esa mano. Asumiendo que las poblaciones son dicotómicas y que las observaciones son aleatorias e independientes, conteste :

a) ¿ Podríamos afirmar que a nivel poblacional existen diferencias en el número de malformaciones entre los hijos de madres expuestas a esa sustancia química y los hijos de las no expuestas ?. Formule las hipótesis nula y alternativa, contraste la hipótesis nula tomando un α = 0'05, e interprete el resultado.

b) Calcule el intervalo de confianza para la diferencia entre los parámetros poblacionales e interprete el resultado.

a) Contraste sobre igualdad de proporciones.

Hipótesis :

H_0 : $\pi_1 = \pi_2$ ($\pi_1 - \pi_2 = 0$) **Muestras independientes**

H_1 : $\pi_1 \neq \pi_2$ ($\pi_1 - \pi_2 \neq 0$)

Función de decisión :

$$z_\alpha = \frac{p_1 - p_2}{D} \quad \text{siendo} \quad D = \sqrt{\left(\frac{n_1.p_1 + n_2.p_2}{n_1+n_2}\right).\left(1 - \frac{n_1.p_1 + n_2.p_2}{n_1+n_2}\right).\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Región de aceptación :

Para un nivel de significación α = 0'05, la tabla normal proporciona los valores $z_\alpha = \pm 1'96$. Es decir, aceptaremos que no existen diferencias significativas si el valor de la función de decisión está en el intervalo **(-1'96 , 1'96)**.

$p_1 = 0'2$; $p_2 = 0'05$; $n_1 = n_2 = 90$; $n_1.p_1 = 18$; $n_2.p_2 = 4'5$

$$D = \sqrt{\left(\frac{18+4'5}{90+90}\right).\left(1 - \frac{18+4'5}{90+90}\right).\left(\frac{1}{90}+\frac{1}{90}\right)} = 0'0493$$

$$z_\alpha = \frac{0'2 - 0'05}{0'0493} = 3'0426 \quad \notin \quad (-1'96\,,\,1'96).$$ Rechazamos la hipótesis nula.

A nivel poblacional, existen diferencias en el número de malformaciones entre los hijos de madres expuestas a la sustancia química y los hijos de las no expuestas.

b) Intervalo de confianza :

$$(p_1 - p_2) \pm z_\alpha . D = (0'2 - 0'05) \pm 1'96\,.\,0'0493 \qquad \Rightarrow \qquad (0'053\,,\,0'247)$$

Con un margen máximo de error del 5% podemos afirmar que, a nivel poblacional, la diferencia en el número de malformaciones entre los hijos de madres expuestas a la sustancia química y los hijos de las no expuestas, está comprendida entre el 5'3% y el 24'7%.

22

En la enseñanza de idiomas se siguen dos métodos; uno tradicional y otro basado en el empleo continuado de sistemas audiovisuales. Para comparar la eficacia de ambos, se seleccionaron al azar 7 alumnos que recibieron enseñanza con el método tradicional y otros 7 con el segundo método. Aplicada una prueba común, se obtuvieron las puntuaciones siguientes :

Método 1	**7**	**5**	**3**	**5**	**2**	**6**	**4**
Método 2	**6**	**8**	**5**	**9**	**7**	**8**	**6**

Utilizando un nivel de significación 0'01 y suponiendo que las distribuciones son normales :

a) Aplique una técnica paramétrica para contrastar la hipótesis de que las varianzas de las puntuaciones de la prueba en las dos poblaciones de alumnos, definidas en función del método de enseñanza seguido, son iguales.

b) Contraste la hipótesis anterior, suponiendo que sabemos que la puntuación poblacional media que ofrece la prueba en alumnos que siguen el método tradicional es de 5'5 puntos, siendo de 6'8 puntos la obtenida en el segundo colectivo.

a) **Comparación de dos varianzas para muestras independientes.**

Hipótesis : (Contraste bilateral)

$$H_0: \quad \sigma_1^2 = \sigma_2^2 \qquad \left(\frac{\sigma_1^2}{\sigma_2^2} = 1\right) \qquad H_1: \quad \sigma_1^2 \neq \sigma_2^2 \qquad \left(\frac{\sigma_1^2}{\sigma_2^2} \neq 1\right)$$

Parámetros muestrales :

Calculamos las medias, varianzas y cuasivarianzas a partir de las siete observaciones de cada muestra. Estos cálculos proporcionan los resultados siguientes :

$$\bar{x}_1 = 4'57 \qquad s_1^2 = 2'53 \qquad s_1'^2 = 2'95$$
$$\bar{x}_2 = 7 \qquad s_2^2 = 1'71 \qquad s_2'^2 = 2$$

Función de decisión :

$$F_{n_1-1,n_2-1,\alpha} = \frac{s_1'^2}{s_2'^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} = \frac{2'95}{2} .1 = 1'475$$

Correspondiendo a una F de Snedecor con 6 y 6 (n_1-1 y n_2-1) grados de libertad, la tabla relativa a

$\alpha/2 = 0'01/2 =$ **0'005** (contraste bilateral)

proporciona los valores tabulados (límites del intervalo de aceptación de la hipótesis nula) :

$$F_{6,6,0'005} = 11'07 \qquad F_{6,6,1-0'005} = F_{6,6,0'995} = \frac{1}{F_{6,6,0'005}} = \frac{1}{11'07} = 0'09$$

Al pertenecer 1'475 al intervalo (0'09 , 11'07), aceptamos la hipótesis de igualdad de varianzas.

b) **Comparación de dos varianzas para muestras independientes y <u>medias poblacionales conocidas</u>.**

Función de decisión :

$$F_{n_1-1,n_2-1,\alpha} = \frac{n_2}{n_1} \cdot \frac{\sum\left(x_1-\mu_1\right)^2}{\sum\left(x_2-\mu_2\right)^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} = \frac{s_1^2 + \left(\bar{x}_1-\mu_1\right)^2}{s_2^2 + \left(\bar{x}_2-\mu_2\right)^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} =$$
$$= \frac{2'53 + (4'57 - 5'5)^2}{1'71 + (7 - 6'8)^2} .1 = 1'9399$$

Dado que 1'9399 pertenece al intervalo (0'09 , 11'07), aceptamos la hipótesis de igualdad de dispersiones (varianzas).

EJERCICIOS PROPUESTOS (Decisión)

1

En la aplicación de una prueba sobre velocidad de respuesta ante estímulos sensoriales, se mantiene que la velocidad media es distinta en hombres y mujeres pero con idéntica dispersión.
Aplicada la prueba a un grupo de 9 hombres y 13 mujeres se calcularon sus cuasivarianzas, siendo 140 y 200 respectivamente.
Con un nivel de significación del 5%, ¿ puede mantenerse que la variabilidad de las velocidades de respuesta es la misma en los dos colectivos ?.

2

Se sabe, por experiencias de cursos anteriores, que el número medio de faltas de ortografía en ejercicios de comentarios sobre textos filosóficos es de 5 faltas.
Calificado un grupo aleatorio de 82 alumnos, se observó que el número medio de faltas era de 2'8 con desviación típica 1'3.
¿ Puede admitirse que el promedio de faltas ortográficas ha disminuido, con una significación del 1% ?.

3

Un psicólogo social realiza un estudio sobre el racismo, manteniendo la creencia de que el comportamiento en zonas urbanas y rurales es distinto. Sus experiencias en ambos grupos le llevan a concluir que entre los habitantes de núcleos urbanos hay un 20% más de personas con sentimientos racistas que en zonas rurales.
Con un nivel de significación $\alpha = 0'05$, justifique si es aceptable la afirmación del psicólogo, partiendo de las observaciones muestrales siguientes :
- de un grupo de 40 habitantes de zonas urbanas, 17 se consideran racistas
- de otro de 32 habitantes en zonas rurales, pueden considerarse racistas a tan sólo 5.

4

En la Escuela Universitaria de Formación del Profesorado se imparte la asignatura de Informática, común a las especialidades de Ciencias y Filología.
Sabiendo que las calificaciones de ambos grupos de alumnos están igualmente dispersas y que en un grupo de 12 de la especialidad de Ciencias se obtuvo una nota media de 6'8 puntos con varianza 4 y, en otro de 10 de Filología la media fue de 4'5 puntos con varianza 3'6, ¿ puede afirmarse que existe una diferencia significativa entre las calificaciones ?.
Realice el correspondiente contraste para los niveles de significación del 5% y del 1%.

5

Para el contraste de la hipótesis de que la media poblacional es igual a 20 frente a la alternativa de que su valor es 28, se selecciona una muestra aleatoria de tamaño 15 en la que se calcula un promedio 25 con desviación típica 4'1.
a) Con una significación máxima del 5%, contraste la hipótesis planteada.
b) Determine la probabilidad de aceptar la hipótesis nula en el supuesto de no ser cierta.
c) ¿ Qué tamaño muestral debemos tomar para que las probabilidades de los errores de tipo I y II sean 0'002 y 0'005 , respectivamente ?.

6

Partiendo de muestras aleatorias de 50 individuos se establece como criterio de aceptación del promedio poblacional $\mu = 5$ frente al alternativo $\mu = 6$ que , para ello, la media muestral observada sea inferior a 5'6.
Si se admite que la varianza poblacional es igual a 2 :
a) Determine las probabilidades de los errores de tipo I y II.
b) Para que el nivel de significación sea del 1%, ¿ qué tamaño muestral debe tomarse ?.

SOLUCIONES DE LOS EJERCICIOS PROPUESTOS

1

Cociente de varianzas para observaciones independientes.

$H_0 :\ \sigma_1^2 / \sigma_2^2 = 1$; $H_1 :\ \sigma_1^2 / \sigma_2^2 \neq 1$

$s'^2_1 = 160$; $s'^2_2 = 218'18$; $F = 0'733$

$F_{8,12,0'025} = 3'51$; $F_{8,12,0'975} = 1 / F_{12,8,0'025} = 1 / 4'20 = 0'2381$

Se admite la igualdad de dispersiones al ser $0'2381 < 0'733 < 3'51$

2

Media con varianza de la población desconocida y muestra grande.

$H_0 :\ \mu \geq 5$; $H_1 :\ \mu < 5$

$z = -15'231$; $z_\alpha = -2'33$; El promedio ha disminuido.

3

Diferencia de proporciones con $\pi_1 - \pi_2 \neq 0$.

$H_0 :\ \pi_1 - \pi_2 = 0'2$; $H_1 :\ \pi_1 - \pi_2 \neq 0'2$

$p_1 = 0'425$; $p_2 = 0'15625$

$D = 0'461$; $z = 0'1491$; $z_{\alpha/2} = \pm 1'96$; Se admite la hipótesis del psicólogo.

4

Diferencia de medias con varianzas poblacionales desconocidas pero iguales (muestras pequeñas).

$H_0 :\ \mu_1 - \mu_2 = 0$; $H_1 :\ \mu_1 - \mu_2 \neq 0$

$t = 2'6211$

$\alpha = 0'05$; $t_{20,0'05} = \pm 2'0860$; Existe una diferencia significativa entre las calificaciones.

$\alpha = 0'01$; $t_{20,0'01} = \pm 2'8453$; No hay diferencia significativa entre las calificaciones.

5

Media con varianza de la población desconocida y muestra pequeña.

$H_0 :\ \mu = 20$; $H_1 :\ \mu = 28$

a) $t = 4'56$; $t_{14,0'05} = 1'7613$; Se rechaza la hipótesis nula $\mu = 20$.

b) inferior a 0’005

c) Estimación puntual de la varianza de la población = Cuasivarianza muestral = 18'011

$z_\alpha = 2'88$; $z_\beta = 2'58$; $n = 8'39 \cong 8$.

6

Media con varianza poblacional conocida.

$H_0 :\ \mu = 5$; $H_1 :\ \mu = 6$

a) $\alpha = Pr(z > 2'97) = 0'00149$; $\beta = Pr(z \leq -1'98) = 0'02385$

b) Para el valor β calculado en a) : $z_\alpha = 2'33$; $z_\beta = 1'98$; $n = 37'15 \cong 37$.

Independientemente del valor β : $n = 30'16 \cong 30$.