Universidad Autónoma de Madrid

Gestión Aeronáutica: Estadística Teórica
Facultad Ciencias Económicas y Empresariales
Departamento de Economía Aplicada
Profesor: Santiago de la Fuente Fernández

INTERVALOS DE CONFIANZA

INTERVALOS DE CONFIANZA

a) Intervalo de confianza para la media μ de una distribución normal $N(\mu, \sigma)$ de varianza conocida σ^2

$$I_{1-\alpha}(\mu) = \left[\bar{x} \ \pm \ z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

b) Intervalo de confianza para la media μ de una distribución normal $N(\mu, \sigma)$ de varianza desconocida σ^2

Muestras grandes $n > 30 \quad \mapsto \quad I_{1-\alpha}(\mu) = \left[\bar{x} \ \pm \ z_{\alpha/2} \dfrac{\hat{s}}{\sqrt{n}} \right]$

Muestras pequeñas $n \leq 30 \quad \mapsto \quad I_{1-\alpha}(\mu) = \left[\bar{x} \ \pm \ t_{\alpha/2;\,(n-1)} \dfrac{\hat{s}}{\sqrt{n}} \right]$

c) Intervalo de confianza para la varianza σ^2 de una distribución normal

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

d) Intervalo de confianza para la diferencia de medias de dos distribuciones normales

- Las varianzas poblaciones σ_1^2 y σ_2^2 son conocidas

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

En todos los intervalos de confianza $\hat{s}^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$ es la cuasivarianza muestral.

- Las varianzas poblaciones σ_1^2 y σ_2^2 son desconocidas:

\- Caso en que la suma $(n_1 + n_2) > 30$ con $n_1 \simeq n_2$

$$I_{1-\alpha}(\mu_1 - \mu_2) = \left[(\bar{x} - \bar{y}) \ \pm \ z_{\alpha/2} \sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}} \right]$$

- Caso en que los tamaños muestrales son pequeños $(n_1+n_2)\leq 30$ y las varianzas son desconocidas, pero iguales ($\sigma_1^2=\sigma_2^2$):

$$I_{1-\alpha}(\mu_1-\mu_2)=\left[(\bar{x}-\bar{y})\ \pm\ t_{\alpha/2;\,(n_1+n_2-2)}\,.\,\hat{s}_p\,.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]$$

$\hat{s}_p^2$ es la media ponderada de las cuasivarianzas muestrales: $\hat{s}_p^2=\dfrac{(n_1-1)\hat{s}_1^2+(n_2-1)\hat{s}_2^2}{n_1+n_2-2}$

- Caso en que los tamaños muestrales son pequeños $(n_1+n_2)\leq 30$ y las varianzas son desconocidas y distintas ($\sigma_1^2\neq\sigma_2^2$):

$$I_{1-\alpha}(\mu_1-\mu_2)=\left[(\bar{x}-\bar{y})\ \pm\ t_{\alpha/2;\,f}\sqrt{\frac{\hat{s}_1^2}{n_1}+\frac{\hat{s}_2^2}{n_2}}\right]$$

f es la aproximación de Welch: $f=\dfrac{\left(\dfrac{\hat{s}_1^2}{n_1}+\dfrac{\hat{s}_2^2}{n_2}\right)^2}{\dfrac{(\hat{s}_1^2/n_1)^2}{n_1+1}+\dfrac{(\hat{s}_2^2/n_2)^2}{n_2+1}}-2$

Cuando el intervalo cubre el 0 no hay diferencia significativa entre las medias poblacionales.

e) Intervalo de confianza para la razón de varianzas de dos poblaciones normales

$$I_{1-\alpha}(\sigma_2^2/\sigma_1^2)=\left[\frac{\hat{s}_1^2/\hat{s}_2^2}{F_{\alpha/2\,;\,(n_1-1),\,(n_2-1)}}\,;\ \frac{\hat{s}_1^2/\hat{s}_2^2}{F_{(1-\alpha/2);\,(n_1-1),\,(n_2-1)}}\right]$$

Cuando el intervalo cubre el 1 no hay diferencia significativa entre las varianzas poblacionales.

Hay que considerar la relación: $F_{\alpha\,;\,n_1,\,n_2}=\dfrac{1}{F_{(1-\alpha);\,n_2,\,n_1}}$

f) Intervalo de confianza para el parámetro p de una distribución binomial de parámetros n, p, B(n, p)

$$I_{1-\alpha}(p)=\left[\hat{p}\ \pm\ z_{\alpha/2}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}\right]$$

g) Intervalo de confianza para la diferencia de parámetros $(p_1 - p_2)$ de dos distribuciones binomiales

$$I_{1-\alpha}(p_1 - p_2) = \left[(\hat{p}_1 - \hat{p}_2) \ \pm \ z_{\alpha/2} \sqrt{\frac{\hat{p}_1 (1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2 (1-\hat{p}_2)}{n_2}} \right]$$

h) Intervalo de confianza para el parámetro λ de una distribución de Poisson

$$I_{1-\alpha}(\lambda) = \left[\hat{\lambda} \ \pm \ z_{\alpha/2} \sqrt{\frac{\hat{\lambda}}{n}} \right]$$

i) Intervalo de confianza para la diferencia de datos apareados

- Para muestras grandes $n > 30$

$$I = \left[\overline{d} \ \pm \ z_{\alpha/2} \frac{\hat{s}_d}{\sqrt{n}} \right]$$

$$d_i = \xi_i - \eta_i \qquad \overline{d} = \frac{1}{n} \sum_{i=1}^{n} d_i \qquad \hat{s}_d^2 = \frac{1}{n-1} \sum_{i=1}^{n} (d_i - \overline{d})^2$$

- Para muestras pequeñas $n \leq 30$ $\quad I = \left[\overline{d} \ \pm \ t_{\alpha/2,\,(n-1)} \frac{\hat{s}_d}{\sqrt{n}} \right]$

CÁLCULO DE INTERVALOS DE CONFIANZA PARA LA MEDIA CON DESVIACIÓN TÍPICA POBLACIONAL CONOCIDA Y DESCONOCIDA.

1.- El peso (en gramos) de las cajas de cereales de una determinada marca sigue una distribución $N(\mu, 5)$. Se han tomado los pesos de 16 cajas seleccionadas aleatoriamente, y los resultados obtenidos han sido:

506, 508, 499, 503, 504, 510, 497, 512, 514, 505, 493, 496, 506, 502, 509, 496.

a) Obtener los intervalos de confianza del 90%, 95% y 99% para la media poblacional.

b) Determinar cuál sería el tamaño muestral necesario para conseguir, con un 95% de confianza, un intervalo de longitud igual a 2 gramos.

c) Suponiendo ahora que σ es desconocida, calcular los intervalos de confianza para la media al 90%, 95% y 99%.

Solución:

a) Se trata de construir un intervalo de confianza para la media poblacional μ de varianza conocida $\sigma^2 = 25$. El intervalo de confianza de nivel $1-\alpha$ viene dado por:

$$I_{1-\alpha}(\mu) = \Big[\overbrace{\overline{x}}^{\text{media muestral}} \pm \overbrace{z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}}^{\text{Error muestral}} \Big] \quad \begin{cases} L_{1-\alpha} = 2\, z_{\alpha/2}\,\dfrac{\sigma}{\sqrt{n}} \;\mapsto\; n = \left(\dfrac{2\, z_{\alpha/2}\,\sigma}{\text{longitud}}\right)^2 & L = \text{longitud o amplitud} \\ \text{Error muestral} = z_{\alpha/2}\,\dfrac{\sigma}{\sqrt{n}} \end{cases}$$

$$\overline{x} = \frac{\sum_{i=1}^{16} x_i}{16} = 503,75 \qquad \begin{cases} 1-\alpha = 0,90 & \alpha = 0,10 & \alpha/2 = 0,05 & z_{\alpha/2} = 1,645 \\ 1-\alpha = 0,95 & \alpha = 0,05 & \alpha/2 = 0,025 & z_{\alpha/2} = 1,96 \\ 1-\alpha = 0,99 & \alpha = 0,01 & \alpha/2 = 0,005 & z_{\alpha/2} = 2,575 \end{cases}$$

Los intervalos de confianza solicitados serán:

$$I_{0,90}(\mu) = \left[503,75 \pm 1,645\,\frac{5}{\sqrt{16}}\right] = \left[503,75 - 1,645\,\frac{5}{\sqrt{16}}\,,\, 503,75 + 1,645\,\frac{5}{\sqrt{16}}\right]$$
$$I_{0,90}(\mu) = [\,501,69\,;\; 505,81\,] \;\equiv\; P[\,501,69 \le \mu \le 505,81\,] = 0,90 = 1-\alpha$$

$$I_{0,95}(\mu) = \left[503,75 \pm 1,96\,\frac{5}{\sqrt{16}}\right] = \left[503,75 - 1,96\,\frac{5}{\sqrt{16}}\,,\, 503,75 + 1,96\,\frac{5}{\sqrt{16}}\right]$$
$$I_{0,95}(\mu) = [\,501,30\,;\; 506,20\,] \;\equiv\; P[\,501,30 \le \mu \le 506,20\,] = 0,95 = 1-\alpha$$

$$I_{0,99}(\mu) = \left[503,75 \pm 2,575\,\frac{5}{\sqrt{16}}\right] = \left[503,75 - 2,575\,\frac{5}{\sqrt{16}}\,,\, 503,75 + 2,575\,\frac{5}{\sqrt{16}}\right]$$
$$I_{0,99}(\mu) = [\,500,53\,;\; 506,97\,] \;\equiv\; P[\,500,53 \le \mu \le 506,97\,] = 0,99 = 1-\alpha$$

La longitud de cada uno de los intervalos de confianza:

$$\left.\begin{array}{l} L_{0,90}(\mu) = 505,81 - 501,69 = 4,12 \\ L_{0,95}(\mu) = 506,20 - 501,30 = 4,9 \\ L_{0,99}(\mu) = 506,97 - 500,53 = 6,44 \end{array}\right]$$

El primer intervalo de confianza es de menor longitud, y, por tanto, podría parecer de más preciso, recordando que su nivel de confianza también es menor.

b) La amplitud o longitud vendrá dado por la fórmula: $I_{1-\alpha}(\mu) = \left[\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$

$$\begin{pmatrix} \text{amplitud o} \\ \text{longitud} \end{pmatrix} = \left(\bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) - \left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 2\, z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \mapsto n = \left(\frac{2\, z_{\alpha/2}\, \sigma}{\text{amplitud}} \right)^2$$

siendo, $n = \left(\frac{2.1{,}96.\,5}{2} \right)^2 \approx 96$ cajas de cereales

c) Se trata de construir un intervalo de confianza para la media poblacional μ de varianza poblacional desconocida, con muestras pequeñas ($n \leq 30$). El intervalo de confianza de nivel $(1-\alpha)$, viene dado por:

$$I_{1-\alpha}(\mu) = \left[\bar{x} \pm t_{(\alpha/2),n-1} \frac{s_x}{\sqrt{n}} \right] \begin{cases} 1-\alpha = 0{,}90 & \alpha = 0{,}10 & t_{0,05;15} = 1{,}753 \\ 1-\alpha = 0{,}95 & \alpha = 0{,}05 & t_{0,025;15} = 2{,}131 \\ 1-\alpha = 0{,}99 & \alpha = 0{,}01 & t_{0,005;15} = 2{,}947 \end{cases}$$

cuasivarianza muestral: $s_x^2 = \frac{\sum_{i=1}^{16} (x_i - \bar{x})^2}{15} = 36{,}037 \mapsto s_x \approx 6$ cuasidesviación típica

Los intervalos de confianza solicitados serán:

$$I_{0,90}(\mu) = \left[503{,}75 \pm 1{,}753 \frac{6}{\sqrt{16}} \right] = \left[503{,}75 - 1{,}753 \frac{6}{\sqrt{16}} , 503{,}75 + 1{,}753 \frac{6}{\sqrt{16}} \right]$$

$$I_{0,90}(\mu) = [\,501{,}12 ;\ 506{,}38\,] \equiv P[\,501{,}12 \leq \mu \leq 506{,}38\,] = 0{,}90 = 1-\alpha$$

$$I_{0,95}(\mu) = \left[503{,}75 \pm 2{,}131 \frac{6}{\sqrt{16}} \right] = \left[503{,}75 - 2{,}131 \frac{6}{\sqrt{16}} , 503{,}75 + 2{,}131 \frac{6}{\sqrt{16}} \right]$$

$$I_{0,95}(\mu) = [\,500{,}55 ;\ 506{,}95\,] \equiv P[\,500{,}55 \leq \mu \leq 506{,}95\,] = 0{,}95 = 1-\alpha$$

$$I_{0,99}(\mu) = \left[503{,}75 \pm 2{,}947 \frac{6}{\sqrt{16}} \right] = \left[503{,}75 - 2{,}947 \frac{6}{\sqrt{16}} , 503{,}75 + 2{,}947 \frac{6}{\sqrt{16}} \right]$$

$$I_{0,99}(\mu) = [\,499{,}33 ;\ 508{,}17\,] \equiv P[\,499{,}33 \leq \mu \leq 508{,}17\,] = 0{,}99 = 1-\alpha$$

Señalar que a mayor nivel de confianza $(1-\alpha)$ mayor es la amplitud del intervalo, y, en consecuencia, los intervalos de confianza son mayores.

CÁLCULO DE INTERVALOS DE CONFIANZA PARA LA MEDIA CON DESVIACIÓN TÍPICA POBLACIONAL CONOCIDA Y DESCONOCIDA.

2.- Una muestra aleatoria extraída de una población normal de varianza 100, presenta una media muestral $\bar{x} = 160$. Con una muestra de tamaño 144, se pide:

a) Calcular un intervalo de confianza del 95 por ciento para la media poblacional.

b) Calcular un intervalo de confianza del 90 por ciento para la media poblacional.

c) Comparar ambos intervalos, desde el punto de vista de la información que generan.

d) Si se quiere tener una confianza del 95 por ciento de que su estimación se encuentra a una distancia de 1,2 cm más o menos de la verdadera media poblacional, ¿cuántas observaciones adicionales deben tomarse?

Solución:

a) Hay que construir un intervalo de confianza para la media poblacional μ de varianza conocida $\sigma^2 = 100$. El intervalo de confianza de nivel $1-\alpha$, viene dado por:

$$I_{1-\alpha}(\mu) = \Big[\overbrace{\bar{x}}^{\text{media muestral}} \pm \overbrace{z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}}^{\text{Error muestral}} \Big] \begin{cases} L_{1-\alpha} = 2\, z_{\alpha/2}\, \dfrac{\sigma}{\sqrt{n}} \quad \mapsto \quad n = \left(\dfrac{2\, z_{\alpha/2}\,\sigma}{\text{longitud}}\right)^2 \quad L = \text{longitud o amplitud} \\ \text{Error muestral} = z_{\alpha/2}\, \dfrac{\sigma}{\sqrt{n}} \end{cases}$$

se tiene que:

$$1-\alpha = 0{,}95 \quad \alpha = 0{,}05 \quad \alpha/2 = 0{,}025 \qquad z_{\alpha/2} = 1{,}96$$
$$\bar{x} = 160 \qquad \sigma^2 = 100 \quad \sigma = 10 \qquad n = 144$$

Intervalo de confianza: $I_{0,95}(\mu) = \left[160 - 1{,}96\,\dfrac{10}{12}\ ;\ 160 + 1{,}96\,\dfrac{10}{12}\right] = [158{,}37\ ;\ 161{,}63]$

b) Es análoga su construcción; la única variación es el nivel de confianza:

$$1-\alpha = 0{,}90 \quad \alpha = 0{,}10 \quad \alpha/2 = 0{,}05 \qquad z_{\alpha/2} = 1{,}645$$

con lo cual, $I_{0,90}(\mu) = \left[160 - 1{,}645\,\dfrac{10}{12}\ ;\ 160 + 1{,}645\,\dfrac{10}{12}\right] = [158{,}63\ ;\ 161{,}37]$

c) Calculando la longitud de cada uno de los dos intervalos de confianza:

$$L_{0,95} = 161{,}63 - 158{,}37 = 3{,}26 \qquad L_{0,90} = 161{,}37 - 158{,}63 = 2{,}74$$

El segundo intervalo de confianza es de longitud menor, y, por tanto, podría parecer más preciso, pero no olvidemos que su nivel de confianza es también menor (el 90 por 100 frente al 95 por ciento del primer intervalo).

d) El error absoluto que se quiere cometer es de 1,2, aplicando la fórmula para la determinación de la muestra a un nivel de confianza del 95 por 100, se tiene:

$$I_{1-\alpha}(\mu) = [\ \overbrace{\bar{x}}^{\text{media muestral}} \pm \overbrace{z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}}^{\text{Error muestral}}\ ] \quad \mapsto \quad \in = z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} \Rightarrow n = \left(\frac{z_{\alpha/2}\,\sigma}{\in}\right)^2$$

$$n = \left(\frac{1{,}96\,.\,10}{1{,}2}\right)^2 \approx 267\cdot$$

En consecuencia, se debería tomar una muestra adicional de 123 elementos $(267 - 144 = 123)$.

CÁLCULO DE INTERVALOS DE CONFIANZA PARA LA MEDIA Y LA VARIANZA CON PARÁMETROS POBLACIONALES DESCONOCIDOS.

3.- La afluencia de visitantes al parque de Monfragüe durante un mes, medida a través de una muestra aleatoria durante 10 días elegidos aleatoriamente, han sido los siguientes:

682, 553, 555, 666, 657, 649, 522, 568, 700, 552

Suponiendo que los niveles de afluencia siguen una distribución normal, y que la desviación típica muestral es de 56,99.

a) Se podría afirmar, con un 95 por ciento de confianza, que la afluencia media al parque es de 600 personas al mes.

b) Los adjudicatarios de la explotación al parque, en negociaciones con la Junta de Extremadura, afirmaron que la afluencia media era constante y que la dispersión sería de unas 15 personas. ¿Queda esta afirmación probada con los datos disponibles con un 95% de confianza?

Solución:

a) Se trata intervalo de confianza para la media μ de una distribución normal de varianza poblacional desconocida σ^2 siendo la muestra pequeña $n \leq 30$

$$I_{1-\alpha}(\mu) = \left[\bar{x} \pm t_{(\alpha/2),n-1}\,\frac{s_x}{\sqrt{n}}\right] = 1-\alpha \begin{cases} n\sigma_x^2 = (n-1)s_x^2 \Rightarrow s_x^2 = \dfrac{n\sigma_x^2}{(n-1)} \\ s_x^2 = \dfrac{10\,.\,56{,}99^2}{9} = 3608{,}73 \mapsto s_x = \sqrt{3608{,}73} = 60{,}07 \\ 1-\alpha = 0{,}95 \quad \alpha = 0{,}05 \quad \alpha/2 = 0{,}025 \quad t_{\alpha/2;(n-1)} = t_{0{,}025;9} = 2{,}262 \end{cases}$$

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = 610{,}04 \qquad s_x = 60{,}07$$

$$I_{0,95}(\mu) = \left[610{,}04 \pm 2{,}262\,\frac{60{,}07}{\sqrt{10}}\right] = \left[610{,}04 - 2{,}262\,\frac{60{,}07}{\sqrt{10}}\,,\,610{,}04 + 2{,}262\,\frac{60{,}07}{\sqrt{10}}\right]$$

$$I_{0,95}(\mu) = [\,567{,}07\,;\ 653{,}01\,] \equiv P[\,567{,}07 \leq \mu \leq 653{,}01\,] = 0{,}95 = 1-\alpha$$

Como $567{,}07 \leq 600 \leq 653{,}01$ se puede afirmar que con un 95 por ciento de confianza la afluencia media es de 600 personas al mes.

b) Intervalo de confianza para la varianza σ^2 de una distribución normal:

$$I_{1-\alpha}(\sigma^2) = \left[\frac{(n-1)s_x^2}{\chi^2_{\alpha/2;(n-1)}} \; ; \; \frac{(n-1)s_x^2}{\chi^2_{1-\alpha/2;(n-1)}} \right] \quad \begin{cases} s_x^2 = 3608,73 \\ 1-\alpha = 0,95 \quad \alpha/2 = 0,025 \quad \chi^2_{\alpha/2;(n-1)} = \chi^2_{0,025;9} = 19,023 \\ \chi^2_{1-\alpha/2;(n-1)} = \chi^2_{0,975;9} = 2,70 \end{cases}$$

$$I_{0,95}(\sigma^2) = \left[\frac{9.(3608,73)}{19,023} \; ; \; \frac{9.(3608,73)}{2,70} \right] = \left[1707,33 \; ; \; 12029,1 \right]$$

$$I_{0,95}(\sigma^2) = \left[1707,33 \; ; \; 12029,1 \right] \equiv P\left[1707,33 \leq \sigma^2 \leq 12029,1 \right] = 0,95 = 1-\alpha$$

$$I_{0,95}(\sigma) = \left[\sqrt{\frac{9.(3608,73)}{19,023}} \; ; \; \sqrt{\frac{9.(3608,73)}{2,70}} \right] = \left[\sqrt{1707,33} \; ; \; \sqrt{12029,1} \right] = \left[41,32 \; ; \; 109,68 \right]$$

$$I_{0,95}(\sigma) = \left[41,32 \; ; \; 109,68 \right] \equiv P\left[41,32 \leq \sigma \leq 109,68 \right] = 0,95 = 1-\alpha$$

$15 \notin \left[41,32 \; ; \; 109,68 \right]$. El intervalo de la desviación típica no contiene el valor 15, con lo cual no se puede afirmar con una confianza del 95% que la dispersión de afluencia sea de 15 personas.

CÁLCULO DE UN INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS CON DESVIACIONES TÍPICAS POBLACIONALES CONOCIDAS.

4.- El gasto diario en llamadas telefónicas de dos departamentos X e Y de una misma empresa sigue una distribución normal, con gasto medio desconocido en ambos. Sin embargo, se conocen las desviaciones típicas, que son 100 y 110 céntimos de euro para X e Y, respectivamente. La dirección ha observado que una muestra aleatoria de 20 días, el gasto medio diario en llamadas realizadas por el departamento X ha sido de 1100 céntimos, y de 1400 en el departamento Y. Obtener un intervalo de confianza para la diferencia de gastos medios entre ambos departamentos.

Solución:

La variables aleatorias siguen, respectivamente, las distribuciones normales $N(\mu_1, 100)$ y $N(\mu_2, 110)$. El intervalo de confianza para la diferencia de medias $(\mu_1 - \mu_2)$ con varianzas poblacionales conocidas viene dado por la expresión:

$$I_{1-\alpha}(\mu_1 - \mu_2) = \left[(\bar{x} - \bar{y}) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right] \quad \begin{cases} \sigma_1^2 = 100^2 \quad \sigma_2^2 = 110^2 \\ \bar{x} = 1100 \quad n_1 = 20 \quad \bar{y} = 1400 \quad n_2 = 20 \\ 1-\alpha = 0,90 \quad \alpha/2 = 0,05 \quad z_{\alpha/2} = 1,645 \end{cases}$$

$$I_{0,90}(\mu_1 - \mu_2) = \left[(1100 - 1400) \pm (1,645) \sqrt{\frac{100^2}{20} + \frac{110^2}{20}} \right] = \left[-354,68 \; ; -245,32 \right]$$

El intervalo de confianza no cubre el 0 por 100, lo que indica que existe diferencia significativa en el gasto de llamadas telefónicas. Como el intervalo de confianza es negativo, se deduce que el gasto medio en llamadas telefónicas del departamento Y es superior al del departamento X, con una confianza del 90 por ciento.

CÁLCULO DE UN INTERVALO DE CONFIANZA PARA LA PROPORCIÓN CON APROXIMACIÓN A UNA NORMAL, AL SER LA MUESTRA SUFICIENTEMENTE GRANDE.

5.- Se selecciona una muestra aleatoria de 600 familias, a las que se pregunta si tienen o no ordenador en casa. Contestaron afirmativamente 240 familias. Obtener un intervalo de confianza al nivel del 95% para la proporción real de familias que poseen ordenador en casa.

Solución:

La característica en estudio es dicotómica, hay que construir un intervalo de confianza para el parámetro p (proporción) de la variable aleatoria binomial asociada al estudio de la característica. Como el tamaño de la muestra es suficientemente grande, n = 600, se puede utilizar la aproximación normal.

$$I_{1-\alpha}(p) = \left[\hat{p} \pm z_{(\alpha/2)} \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} \right] \begin{cases} \hat{p} = 240/600 = 0,4 \quad \hat{q} = 1-\hat{p} = 0,6 \quad n = 600 \\ 1-\alpha = 0,95 \quad \alpha = 0,05 \quad \alpha/2 = 0,025 \quad z_{\alpha/2} = z_{0,025} = 1,96 \end{cases}$$

$$I_{0,95}(p) = \left[0,4 \pm (1,96) \sqrt{\frac{0,4\,.\,0,6}{600}} \right] = [\,0,36\ ;\ 0,44\,]$$

$$I_{0,95}(p) = [\,0,36\ ;\ 0,44\,] \equiv P[\,0,36 \leq p \leq 0,44\,] = 0,95 = 1-\alpha$$

Con una confianza del 95% se puede afirmar que las familias poseen ordenador entre el 36% y el 44%.

CÁLCULO DE UN INTERVALO DE CONFIANZA PARA LA PROPORCIÓN Y PARA LA DIFERENCIA DE PROPORCIONES. CÁLCULO DE LA AMPLITUD Y ANÁLISIS DEL ERROR DE ESTIMACIÓN.

6.- Según los dirigentes del partido A, la intención de voto del partido rival B, en Andalucía, es la misma que la que tiene en Madrid. Se realiza una encuesta a 100 personas en Andalucía de los que 25 mostraron su apoyo al partido B, y a otras 100 personas en Madrid de las que 30 se inclinaron por el partido B.

a) Construir un intervalo de confianza del 90% para la proporción de personas que votarían al partido B en Andalucía

b) ¿A cuántas personas habría que encuestar para obtener un margen de error o error de estimación $\pm$ 2%, al nivel de confianza anterior?.

c) Construir un intervalo de confianza al 90% para la diferencia de proporciones en la estimación del voto del partido B en las dos comunidades. ¿Podemos afirmar que los dirigentes del partido A tienen razón?.

Solución:

a) La característica en estudio en ambas comunidades es dicotómica, tenemos que construir un intervalo de confianza para el parámetro p_1 (proporción) de la variable aleatoria binomial asociada al estudio de la característica en la comunidad de Andalucía. Como el tamaño de la muestra es suficientemente grande, n_1 = 100, se puede utilizar la aproximación normal.

$$I_{1-\alpha}(p)=\left[\hat{p}\pm z_{(\alpha/2)}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}\right]\quad \begin{cases}\hat{p}_1 = 25/100 = 0,25 & \hat{q}_1 = 1-\hat{p}_1 = 0,75 \qquad n_1 = 100\\ 1-\alpha = 0,90 \quad \alpha = 0,10 & \alpha/2 = 0,05 \quad z_{\alpha/2} = z_{0,05} = 1,645\end{cases}$$

$$I_{0,90}(p_1)=\left[0,25\pm(1,645)\sqrt{\frac{0,25\,.\,0,75}{100}}\right] = [\,0,179\ ;\ 0,321\,]$$

$$I_{0,90}(p_1) = [\,0,179\ ;\ 0,321\,] \equiv P\left[\,0,179 \le p_1 \le 0,321\,\right] = 0,90 = 1-\alpha$$

En Andalucía la intención de voto del partido B se encuentra entre el 17,9% y 32,1%, con un nivel de confianza del 90%.

b) La amplitud o longitud vendrá dado por la fórmula:

$$I=\left[\begin{pmatrix}\text{proporción}\\ \text{muestral}\end{pmatrix}\pm\begin{pmatrix}\text{error}\\ \text{muestral}\end{pmatrix}\right]\equiv\left[\hat{p}_x + z_{(\alpha/2)}\sqrt{\frac{\hat{p}_x\,.\,\hat{q}_x}{n}}\right] \quad\mapsto\quad \in^2 = \left(z_{\alpha/2}\sqrt{\frac{\hat{p}_x\,.\,\hat{q}_x}{n}}\right)^2$$

de donde, $n=\dfrac{(z_{\alpha/2})^2\,(\hat{p}_x\,.\,\hat{q}_x)}{\in^2}$

El caso más desfavorable será cuando $\hat{p}_x = \hat{q}_x = 0,5$.

Siendo $\in^2 = (\pm 0,02)^2 = 0,0004 \quad\mapsto\quad n=\dfrac{(1,645)^2\,(0,5\,.\,0,5)}{0,0004}\simeq 1691$

c) Nos encontramos ante un intervalo de confianza para la diferencia de parámetros poblacionales $(p_1 - p_2)$ de dos distribuciones binomiales, con el tamaño de las muestras suficientemente grandes, $n_1 = n_2 = 100$, para utilizar la aproximación normal.

$$I_{1-\alpha}(p_1-p_2)=\left[(\hat{p}_1-\hat{p}_2)\pm z_{(\alpha/2)}\sqrt{\frac{\hat{p}_1\,(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2\,(1-\hat{p}_2)}{n_2}}\right]$$

$$\begin{cases}\hat{p}_1 = 25/100 = 0,25 & \hat{q}_1 = 1-\hat{p}_1 = 0,75 \qquad n_1 = 100\\ \hat{p}_2 = 30/100 = 0,3 & \hat{q}_2 = 1-\hat{p}_2 = 0,70 \qquad n_2 = 100\\ 1-\alpha = 0,90 \quad \alpha = 0,10 & \alpha/2 = 0,05 \quad z_{\alpha/2} = z_{0,05} = 1,645\end{cases}$$

$$I_{0,90}(p_1-p_2)=\left[(0,25-0,3)\pm(1,645)\sqrt{\frac{0,25\,.\,0,75}{100}+\frac{0,3\,.\,0,70}{100}}\right]=[-0,153\ ;\ 0,053]$$

El intervalo de confianza cubre el cero, lo que indica que no existe diferencia significativa entre la intención de voto del partido B en ambas comunidades, con lo cual los dirigentes del partido A tienen razón con una fiabilidad del 90%.

CÁLCULO DE UN INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS POBLACIONALES CON DESVIACIONES TÍPICAS POBLACIONALES CONOCIDAS O DESCONOCIDAS.

7.- Un fabricante de televisores está desarrollando un nuevo modelo de televisor en color, y para este fin se pueden utilizar dos tipos de esquemas transistorizados. El fabricante selecciona una muestra de esquemas transistorizados del primer tipo de tamaño 16, y otra del segundo tipo de tamaño 13. Los datos muestrales respecto a la vida media de cada esquema son los siguientes:

$$\bar{x}_1 = 1400 \text{ horas} \qquad s_1 = 30 \text{ horas} \qquad n_1 = 16$$
$$\bar{x}_2 = 1500 \text{ horas} \qquad s_2 = 17 \text{ horas} \qquad n_2 = 13$$

Construir un intervalo de confianza del 90% para la diferencia de vida media de cada tipo de esquema.

Solución:

Sea la variable aleatoria X_1 = 'vida media del primer esquema', que sigue una distribución normal $N(\mu_1, \sigma_1)$. Análogamente, la variable aleatoria X_2 = 'vida media del segundo esquema', sigue una distribución normal $N(\mu_2, \sigma_2)$.

Hay que construir un intervalo de confianza para la diferencia de medias poblacionales $(\mu_1 - \mu_2)$ con varianzas poblacionales desconocidas, y no sabemos si distintas o no, siendo las muestras pequeñas $n_1 + n_2 = 29 < 30$.

Para dilucidar si las varianzas poblacionales desconocidas son o no distintas, construimos primero un intervalo de confianza para el cociente de varianzas (σ_1^2/σ_2^2), de modo que si el intervalo cubre al punto 1 podremos partir de que las varianzas son desconocidas pero iguales.
Para construir un intervalo de confianza para el cociente de varianzas se emplea la fórmula:

$$I_{1-\alpha}(\sigma_1^2/\sigma_2^2) = \left[\frac{s_1^2/s_2^2}{F_{\alpha/2;(n_1-1),(n_2-1)}} \; ; \; \frac{s_1^2/s_2^2}{F_{1-\alpha/2;(n_1-1),(n_2-1)}} \right] \text{ siendo } F_{1-\alpha/2;(n_1-1),(n_2-1)} = \frac{1}{F_{\alpha/2;(n_2-1),(n_1-1)}}$$

$$s_1^2 = 30^2 = 900 \quad n_1 = 16 \quad s_2^2 = 17^2 = 289 \quad n_2 = 13 \quad s_1^2/s_2^2 = 900/289 = 3{,}114$$
$$1-\alpha = 0{,}90 \quad \alpha = 0{,}10 \quad \alpha/2 = 0{,}05$$
$$F_{0,05;\,15,12} = 2{,}6169 \quad F_{0,95;\,15,12} = 1/F_{0,05;\,12,15} = 1/2{,}4753 = 0{,}404$$

de donde,

$$I_{0,90}(\sigma_1^2/\sigma_2^2) = \left[\frac{3{,}114}{2{,}6169} \; ; \; \frac{3{,}114}{0{,}404} \right] = [1{,}19 \; ; 7{,}71]$$

$$I_{0,90}(\sigma_1^2/\sigma_2^2) = [1{,}19 \; ; 7{,}71] \equiv P\left[1{,}19 \le \frac{\sigma_1^2}{\sigma_2^2} \le 7{,}71\right] = 0{,}90 = 1-\alpha$$

El intervalo no cubre el punto uno, y concluimos que las varianzas poblacionales son desconocidas y distintas, con una fiabilidad del 90%.

Nos situamos ante un intervalo de confianza para la diferencia de medias poblacionales $(\mu_1 - \mu_2)$ con varianzas poblacionales desconocidas y distintas o no, con muestras pequeñas $n_1 + n_2 = 29 < 30$.

$$I_{1-\alpha}(\mu_1 - \mu_2) = \left[(\bar{x} - \bar{y}) \pm t_{\alpha/2,f} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right] \begin{cases} t_{\alpha/2,f} \ \text{ donde f es la aproximación de Welch} \\ f = \dfrac{(s_1^2/n_1 + s_2^2/n_2)^2}{\dfrac{(s_1^2/n_1)^2}{n_1+1} + \dfrac{(s_2^2/n_2)^2}{n_2+1}} - 2 \end{cases}$$

siendo,

$s_1^2 = 30^2 = 900 \quad n_1 = 16 \quad s_2^2 = 17^2 = 289 \quad n_2 = 13$

$s_1^2/n_1 = 900/16 = 56,25 \quad s_2^2/n_2 = 289/13 = 22,23$

$(s_1^2/n_1)^2 = 3164,06 \quad (s_2^2/n_2)^2 = 494,17 \quad (s_1^2/n_1)^2/17 = 186,12 \quad (s_2^2/n_2)^2/14 = 13,294$

$(s_1^2/n_1 + s_2^2/n_2)^2 = 6159,11$

$$f = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\dfrac{(s_1^2/n_1)^2}{n_1+1} + \dfrac{(s_2^2/n_2)^2}{n_2+1}} - 2 = \frac{6159,11}{186,12 + 13,294} - 2 = 28,89 \simeq 29$$

$1-\alpha = 0,90 \quad \alpha = 0,10 \quad \alpha/2 = 0,05 \quad t_{\alpha/2,f} = t_{0,05\,;\,29} = 1,699$

$\bar{x}_1 = 1400$ horas $\quad \bar{x}_2 = 1500$ horas

$$I_{0,90}(\mu_1 - \mu_2) = \left[(1400 - 1500) \pm (1,699) \sqrt{\frac{900}{16} + \frac{289}{13}} \right] = [-115,05\,;\,-84,95]$$

El intervalo no cubre el cero, concluyendo que existe diferencia significativa entre la vida media de cada esquema, siendo mayor la vida media del segundo esquema con una fiabilidad del 90%.

CÁLCULO DE UN INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS POBLACIONALES CON DESVIACIONES TÍPICAS POBLACIONALES CONOCIDAS O DESCONOCIDAS.

8.- Un instituto de investigaciones agronómicas siembra, en cinco parcelas diferentes, dos tipos de maíz híbrido. Las producciones en quintales métricos por hectárea son:

	1	2	3	4	5
Híbrido I	90	85	95	76	80
Híbrido II	84	87	90	92	90

a) Construir un intervalo de confianza para el cociente de varianzas con un error de significación de 0,10.

b) Construir un intervalo de confianza del 90% para la diferencia entre las producciones medias.

Solución:

a) Sea la variable aleatoria X_1 = 'producción de maíz del híbrido I', que sigue una distribución normal $N(\mu_1, \sigma_1)$. Análogamente, la variable aleatoria X_2 = 'producción de maíz del híbrido II', sigue una distribución normal $N(\mu_2, \sigma_2)$.

Al construir un intervalo de confianza para el cociente de varianzas podremos concluir si las varianzas poblacionales desconocidas son o no distintas.
De modo que, si el intervalo de confianza para el cociente de varianzas (σ_1^2/σ_2^2) cubre al punto 1 podremos partir de que las varianzas son desconocidas pero iguales.

$$I_{1-\alpha}(\sigma_1^2/\sigma_2^2) = \left[\frac{s_1^2/s_2^2}{F_{\alpha/2;(n_1-1),(n_2-1)}} \; ; \; \frac{s_1^2/s_2^2}{F_{(1-\alpha/2);(n_1-1),(n_2-1)}}\right] \text{ donde } F_{(1-\alpha/2);(n_1-1),(n_2-1)} = \frac{1}{F_{\alpha/2;(n_2-1),(n_1-1)}}$$

En el caso,

$$\bar{x}_1 = 85{,}20 \quad s_1^2 = 57{,}7 \quad n_1 = 5$$
$$\bar{x}_2 = 88{,}6 \quad s_2^2 = 9{,}8 \quad n_2 = 5$$
$$s_1^2/s_2^2 = 57{,}7/9{,}8 = 5{,}89 \quad 1-\alpha = 0{,}90 \quad \alpha = 0{,}10 \quad \alpha/2 = 0{,}05$$
$$F_{0,05\,;\,4,4} = 6{,}3883 \quad F_{0,95\,;\,4,4} = 1/F_{0,05\,;\,4,4} = 1/6{,}3883 = 0{,}1565$$

$$I_{0,90}(\sigma_1^2/\sigma_2^2) = \left[\frac{5{,}89}{6{,}3883} \; ; \; \frac{5{,}89}{0{,}1565}\right] = [\,0{,}92\,;\,37{,}64\,]$$

$$I_{0,90}(\sigma_1^2/\sigma_2^2) = [\,0{,}92\,;\,37{,}64\,] \equiv P\left[0{,}92 \le \frac{\sigma_1^2}{\sigma_2^2} \le 37{,}64\right] = 0{,}90 = 1-\alpha$$

El intervalo cubre el uno, y concluimos que las varianzas poblacionales son desconocidas e iguales, con una fiabilidad del 90%.
b) Nos situamos ante un intervalo de confianza para la diferencia de medias poblacionales $(\mu_1 - \mu_2)$ con varianzas poblacionales desconocidas pero iguales, con muestras pequeñas $n_1 + n_2 = 10 < 30$.

$$I_{1-\alpha}(\mu_1-\mu_2)=\left[(\bar{x}-\bar{y})\pm t_{\alpha/2\,,(n_1+n_2-2)}\; s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]$$

donde, $s_p^2 \equiv$ media ponderada de las cuasivarianzas muestrales:

$$s_p^2=\frac{(n_1-1)\,s_1^2+(n_2-1)\,s_2^2}{n_1+n_2-2} \quad\mapsto\quad s_p^2=\frac{4\,(57,7)+4\,(9,8)}{5+5-2}=33,75 \qquad s_p=5,81$$

$\bar{x}_1=85,20 \quad \bar{x}_2=88,6 \quad n_1=n_2=5$

$1-\alpha=0,90 \quad \alpha=0,10 \quad \alpha/2=0,05 \quad t_{\alpha/2\,,(n_1+n_2-2)}=t_{0,05\,;\,8}=1,860$

$$I_{0,90}(\mu_1-\mu_2)=\left[(85,20-88,6)\;\pm\;(1,860)\,(5,81)\sqrt{\frac{1}{5}+\frac{1}{5}}\right]=\left[\,-10,23\,;\,3,43\,\right]$$

$$I_{0,90}(\mu_1-\mu_2)=\left[\,-10,23\,;\,3,43\,\right]\;\mapsto\; P\left[\,-10,23\le\mu_1-\mu_2\le 3,43\right]=0,90$$

El intervalo de confianza cubre el cero, por lo que no existe diferencia significativa entre las producciones medias, con una fiabilidad del 90%.

CÁLCULO DE UN INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE DATOS APAREADOS.

9.- Un equipo de investigación biológica está interesado en ver si una nueva droga reduce el colesterol en la sangre. Con tal fin toma una muestra de diez pacientes y determina el contenido de colesterol en la sangre antes y después del tratamiento. Los datos muestrales expresados en miligramos por 100 mililitros son los siguientes:

Paciente	1	2	3	4	5	6	7	8	9	10
Antes	217	252	229	200	209	213	215	260	232	216
Después	209	241	230	208	206	211	209	228	224	203

Construir un intervalo de confianza del 95 por 100 para la diferencia del contenido medio de colesterol en la sangre antes y después del tratamiento.

Solución:

Se trata de datos apareados, en los que no existe independencia entre las muestras.
En este caso, como la muestra es pequeña $(n=10<30)$ el intervalo de confianza es:

$$I_{1-\alpha}(\mu_1-\mu_2)=\left[\bar{d}\pm t_{\alpha/2\,,(n-1)}\frac{s_d}{\sqrt{n}}\right] \quad \begin{cases} d_i=x_i-y_i \quad \bar{d}=\dfrac{\sum_{i=1}^{n} d_i}{n} \quad s_d^2=\dfrac{\sum_{i=1}^{n}(d_i-\bar{d})^2}{n-1} \end{cases}$$

donde $\bar{d}$ es la media de las diferencias y s_d la desviación estándar de estas diferencias.

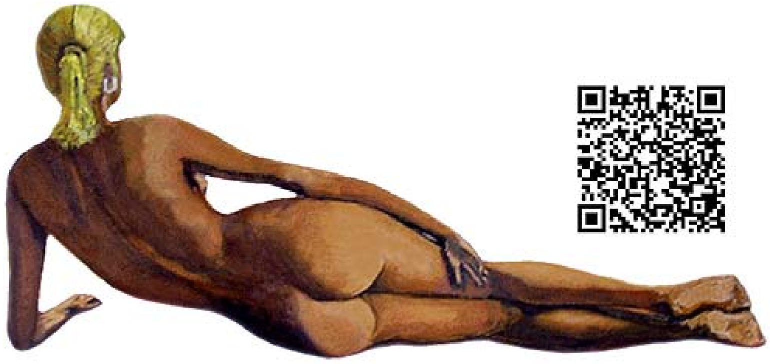
X = 'Antes'	217	252	229	200	209	213	215	260	232	216
Y = 'Después'	209	241	230	208	206	211	209	228	224	203
$d_i = x_i - y_i$	8	11	- 1	- 8	3	2	6	32	8	13

$\overline{d} = 7,40 \quad s_d^2 = 112,1481 \quad s_d = 10,59 \quad n = 10$

$1 - \alpha = 0,95 \quad \alpha = 0,05 \quad \alpha/2 = 0,025 \quad t_{\alpha/2\,;(n-1)} = t_{0,025\,;\,9} = 2,262$

$$I_{0,95}(\mu_1 - \mu_2) = \left[7,40 \pm (2,262)\frac{10,59}{\sqrt{10}} \right] = \left[-0,17\,;\; 14,97 \right]$$

El intervalo abarca el cero, por lo que no existe diferencia significativa en la diferencia del contenido medio del colesterol antes y después del tratamiento, con una fiabilidad del 95%.



UAM
UNIVERSIDAD AUTONOMA
DE MADRID

Asignatura..

Apellidos..

Ejercicio del día..

ESQUEMA INTERVALOS DE CONFIANZA

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

ESQUEMA INTERVALOS DE CONFIANZA

$N(\mu,\sigma)$

$n \quad \overline{x} \quad \sigma_x \quad s_x$

$n\,\sigma_x^2 = (n-1)\,s_x^2$

$$\begin{cases} \overline{x} \sim N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right) \\ \chi^2_{\alpha/2,n-1} = \dfrac{\text{varianza muestral}}{\text{varianza teórica}} = \dfrac{\sigma_x^2}{(\sigma/\sqrt{n})^2} = \dfrac{n\sigma_x^2}{\sigma^2} = \dfrac{(n-1)s_x^2}{\sigma^2} \\ \sigma^2 = \dfrac{(n-1)s_x^2}{\chi^2_{\alpha/2,n-1}} \end{cases}$$

INTERVALOS DE CONFIANZA

MEDIA POBLACIONAL

$$I_{1-\alpha}(\mu) = \left[\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

VARIANZA POBLACIONAL

$$I_{1-\alpha}(\sigma^2) = \left[\frac{(n-1)s_x^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s_x^2}{\chi^2_{1-\alpha/2,n-1}}\right]$$

σ^{*} **desconocida** $N(\mu,\sigma^{*})$

$n \quad \overline{x} \quad \sigma_x \quad s_x$

$n\,\sigma_x^2 = (n-1)\,s_x^2$

$$\frac{\sigma_x}{\sqrt{n-1}} = \frac{s_x}{\sqrt{n}}$$

$$\begin{cases} n > 30 \quad \overline{x} \sim N\left(\mu, \dfrac{s_x}{\sqrt{n}}\right) \\ n \le 30 \quad \overline{x} \sim t_{\alpha/2,n-1} \quad t_{\alpha/2,n-1} = \dfrac{\overline{x}-\mu}{s_x/\sqrt{n}} = \dfrac{\overline{x}-\mu}{\sigma_x/\sqrt{n-1}} \end{cases}$$

INTERVALOS DE CONFIANZA

MEDIA POBLACIONAL ($n > 30$)

$$I_{1-\alpha}(\mu) = \left[\overline{x} \pm z_{\alpha/2}\frac{s_x}{\sqrt{n}}\right]$$

MEDIA POBLACIONAL ($n \le 30$)

$$I_{1-\alpha}(\mu) = \left[\overline{x} \pm t_{\alpha/2,n-1}\frac{s_x}{\sqrt{n}}\right]$$

$X \sim B(n,p) \to N(np, \sqrt{npq})$

$n \quad \hat{p}$

$(X_1, X_2, \cdots, X_n)$

$$\hat{p} = \frac{X}{n} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$$

INTERVALOS DE CONFIANZA PARA EL PARÁMETRO p

$$I_{1-\alpha}(p) = \left[\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right]$$

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

$X \sim B(n_1, p_1) \to N(n_1 p_1, \sqrt{n_1 p_1 q_1})$ $\quad$ $Y \sim B(n_2, p_2) \to N(n_2 p_2, \sqrt{n_2 p_2 q_2})$

$n_1 \ \hat{p}_1$ $\quad - \quad$ $n_2 \ \hat{p}_2$

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right)$$

$$\hat{p}_1 = \frac{X}{n_1} \sim N\left(p_1, \sqrt{\frac{p_1 q_1}{n_1}}\right) \qquad \hat{p}_2 = \frac{Y}{n_2} \sim N\left(p_2, \sqrt{\frac{p_2 q_2}{n_2}}\right)$$

INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE PARÁMETROS POBLACIONALES

$$I_{1-\alpha}(p_1 - p_2) = \left[\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}\right]$$

$N(\mu_1, \sigma_1)$ $\quad$ $N(\mu_2, \sigma_2)$

n_1, $\bar{x} \sim N\left(\mu_1, \sqrt{\frac{\sigma_1}{n_1}}\right)$ $\quad - \quad$ n_2, $\bar{y} \sim N\left(\mu_2, \sqrt{\frac{\sigma_2}{n_2}}\right)$

$$\bar{x} - \bar{y} \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS POBLACIONALES

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

$N(\mu_1, \sigma_1^{\bullet})$ $\quad$ $N(\mu_2, \sigma_2^{\bullet})$

$n_1 \ \hat{s}_1^2$, $\bar{x} \sim N\left(\mu_1, \sqrt{\frac{\hat{s}_1^2}{n_1}}\right)$ $\quad - \quad$ $n_2 \ \hat{s}_2^2$, $\bar{y} \sim N\left(\mu_2, \sqrt{\frac{\hat{s}_2^2}{n_2}}\right)$

$\sigma^{\bullet}$ **desconocida** $n_1 + n_2 > 30$ con $n_1 = n_2$

$$\bar{x} - \bar{y} \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}}\right)$$

INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS POBLACIONALES

$$I_{1-\alpha}(\mu_1 - \mu_2) = \left[\bar{x} - \bar{y} \pm z_{\alpha/2} \sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}}\right]$$

$N(\mu_1, \sigma_1^{\bullet})$ $\quad$ $N(\mu_2, \sigma_2^{\bullet})$

$n_1 \ \hat{s}_1^2$, $\bar{x} \sim N\left(\mu_1, \sqrt{\frac{\hat{s}_1^2}{n_1}}\right)$ $\quad - \quad$ $n_2 \ \hat{s}_2^2$, $\bar{y} \sim N\left(\mu_2, \sqrt{\frac{\hat{s}_2^2}{n_2}}\right)$

$\sigma_1^{\bullet}$ y $\sigma_2^{\bullet}$ **desconocidas,** $n_1 + n_2 \le 30$, con $\sigma_1^{\bullet} = \sigma_2^{\bullet}$

$$\bar{x} - \bar{y} \sim t_{\alpha/2;\,(n_1+n_2-2)}\left(\mu_1 - \mu_2, \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

$$\hat{s}_p^2 = \frac{(n_1 - 1)\hat{s}_1^2 + (n_2 - 1)\hat{s}_2^2}{n_1 + n_2 - 2}$$

INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS POBLACIONALES CON MUESTRAS PEQUEÑAS Y VARIANZAS POBLACIONALES DESCONOCIDAS E IGUALES

$$I_{1-\alpha}(\mu_1 - \mu_2) = \left[(\bar{x} - \bar{y}) \pm t_{\alpha/2,(n_1+n_2-2)}\, \hat{s}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$$

UNIVERSIDAD AUTÓNOMA DE MADRID

Asignatura .... Grupo ....
Apellidos .... Nombre ....
Ejercicio del día ....

$N(\mu_1, \sigma_1^{\bullet})$ $\qquad$ $N(\mu_2, \sigma_2^{\bullet})$

$$\left[\begin{array}{l} n_1 \;\; \hat{s}_1^2 \\ \bar{x} \sim N\left(\mu_1, \sqrt{\dfrac{\hat{s}_1^2}{n_1}}\right) \end{array}\right] - \left[\begin{array}{l} n_2 \;\; \hat{s}_2^2 \\ \bar{y} \sim N\left(\mu_2, \sqrt{\dfrac{\hat{s}_2^2}{n_2}}\right) \end{array}\right] \longrightarrow \bar{x} - \bar{y} \sim t_{\alpha/2;f}\left(\mu_1 - \mu_2, \sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}}\right)$$

$\sigma_1^{\bullet}$ y $\sigma_2^{\bullet}$ desconocidas, $n_1 + n_2 \leq 30$, con $\sigma_1^{\bullet} \neq \sigma_2^{\bullet}$

$$f = \frac{\left(\dfrac{\hat{s}_1^2}{n_1} + \dfrac{\hat{s}_2^2}{n_2}\right)^2}{\dfrac{(\hat{s}_1^2/n_1)^2}{n_1+1} + \dfrac{(\hat{s}_2^2/n_2)^2}{n_2+1}} - 2$$ Aproximación de Welch

INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS POBLACIONALES CON MUESTRAS PEQUEÑAS Y VARIANZAS POBLACIONALES DESCONOCIDAS Y DISTINTAS

$$I_{1-\alpha}(\mu_1 - \mu_2) = \left[(\bar{x} - \bar{y}) \pm t_{\alpha/2;f}\sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}}\right]$$

$N(\mu_1, \sigma_1)$ $\qquad$ $N(\mu_2, \sigma_2)$

$$\left[\begin{array}{l} n_1 \;\; \hat{s}_1^2 \\ \bar{x} \sim N\left(\mu_1, \sqrt{\dfrac{\hat{s}_1^2}{n_1}}\right) \end{array}\right] - \left[\begin{array}{l} n_2 \;\; \hat{s}_2^2 \\ \bar{y} \sim N\left(\mu_2, \sqrt{\dfrac{\hat{s}_2^2}{n_2}}\right) \end{array}\right] \longrightarrow \sigma_1^2/\sigma_1^2 \sim F_{\alpha/2;(n_1-1),(n_2-1)} \quad F_{\alpha/2;n_1,n_2} = \frac{1}{F_{1-\alpha/2;n_2,n_1}}$$

INTERVALO DE CONFIANZA PARA LA RAZÓN DE VARIANZAS DE DOS POBLACIONES NORMALES

$$I_{1-\alpha}(\sigma_1^2/\sigma_1^2) = \left[\frac{\hat{s}_1^2/\hat{s}_2^2}{F_{\alpha/2;(n_1-1),(n_2-1)}};\ \frac{\hat{s}_1^2/\hat{s}_2^2}{F_{1-\alpha/2;(n_1-1),(n_2-1)}}\right]$$

ÚNICAMENTE PARA USO ESCOLAR (IMPRESO EN PAPEL RECICLADO)

Gestión Aeronáutica: Estadística Teórica
Facultad Ciencias Económicas y Empresariales
Departamento de Economía Aplicada
Profesor: Santiago de la Fuente Fernández