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

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

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

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

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 $s_x^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$ 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 $n_1 \simeq n_2$

$$I_{1-\alpha}(\mu_1-\mu_2)=\left[(\bar{x}-\bar{y})\ \pm\ z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right]$$

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

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

s_p^2 es la media ponderada de las cuasivarianzas muestrales: $s_p^2=\dfrac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$

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

$$I_{1-\alpha}(\mu_1-\mu_2)=\left[(\bar{x}-\bar{y})\ \pm\ t_{\alpha/2\,;\,f}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right]$$

f es la aproximación de Welch: $f=\dfrac{\left(\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}\right)^2}{\dfrac{(s_1^2/n_1)^2}{n_1+1}+\dfrac{(s_2^2/n_2)^2}{n_2+1}}-2$

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

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

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

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

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

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 $n > 30$

$$I = \left[\bar{d} \ \pm \ z_{\alpha/2} \frac{s_d}{\sqrt{n}} \right]$$

$$d_i = \xi_i - \eta_i \qquad \bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \qquad s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$$

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

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:

$$I_{1-\alpha}(\mu) = \Big[\underbrace{\overline{x}}_{\text{media muestral}} \pm \underbrace{z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\text{error estimación}}\Big]$$

$$\begin{cases} L_{1-\alpha} = 2\,z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} \;\rightarrow\; n = \left(\dfrac{2\,z_{\alpha/2}\,\sigma}{L_{1-\alpha}}\right)^2 & L = \text{longitud o amplitud} \\ \text{Error estimación} \equiv \in = z_{z_{\alpha/2}}\dfrac{\sigma}{\sqrt{n}} \end{cases}$$

$$\overline{x} = \frac{1}{16}\sum_{i=1}^{16} x_i = 503,75$$

$$\begin{cases} 1-\alpha = 0,90 & \alpha = 0,10 & \alpha/2 = 0,05 & z_{0,05} = 1,645 \\ 1-\alpha = 0,95 & \alpha = 0,05 & \alpha/2 = 0,025 & z_{0,025} = 1,96 \\ 1-\alpha = 0,99 & \alpha = 0,01 & \alpha/2 = 0,005 & z_{0,005} = 2,575 \end{cases}$$

A medida que el nivel de confianza es mayor, aumenta longitud del intervalo.

$$P\left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$$

Los intervalos de confianza solicitados serán:

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

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

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

Longitud de cada uno de los intervalos de confianza:

$$L_{0,90}(\mu) = 505,81 - 501,69 = 4,12$$
$$L_{0,95}(\mu) = 506,20 - 501,30 = 4,9$$
$$L_{0,99}(\mu) = 506,97 - 500,53 = 6,44$$

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

$$\begin{pmatrix} \text{Amplitud o} \\ \text{Longitud} \end{pmatrix} = \left(\overline{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) - \left(\overline{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 2\, z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \;\rightarrow\; n = \left(\frac{2\, z_{\alpha/2}\, \sigma}{\text{Amplitud}} \right)^2$$

siendo $n = \left(\dfrac{2 \times 1,96 \times 5}{2} \right)^2 \approx 96$ **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:

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

Cuasivarianza muestral: $s_x^2 = \dfrac{1}{15} \sum_{i=1}^{16} (x_i - \overline{x})^2 = 36,037 \;\rightarrow\; s_x \approx 6$ **cuasidesviación típica**

Los intervalos de confianza solicitados serán:

$$I_{0,90}(\mu) = \left[503,75 \pm 1,753 \frac{6}{\sqrt{16}} \right] = \left[501,12\,,\ 506,38 \right]$$

$$I_{0,95}(\mu) = \left[503,75 \pm 2,131 \frac{6}{\sqrt{16}} \right] = \left[500,55\,,\ 506,95 \right]$$

$$I_{0,99}(\mu) = \left[503,75 \pm 2,947 \frac{6}{\sqrt{16}} \right] = \left[499,33\,,\ 508,17 \right]$$

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

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

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

b) 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) $n = 144 \quad \overline{x} = 160 \quad \sigma = 10 \quad 1-\alpha = 0,95 \quad \alpha/2 = 0,025 \quad z_{0,025} = 1,96$

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

b) 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:

$$\in=\,z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\;\Rightarrow\;n=\left(\frac{z_{\alpha/2}\,\sigma}{\in}\right)^2\;\;\rightarrow\;\;n=\left(\frac{1,96\,.\,10}{1,2}\right)^2\approx 267$$

Se debería tomar una muestra adicional de $267-144=123$ elementos

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 siendo la muestra pequeña $n\leq 30$

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

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

$$I_{0,95}(\mu)\;=\;\left[610,04\pm 2,262\,\frac{60,07}{\sqrt{10}}\right]\;=\;\left[610,04-2,262\,\frac{60,07}{\sqrt{10}}\,,610,04+2,262\,\frac{60,07}{\sqrt{10}}\right]=\left[\,567,07\,;\,653,01\,\right]$$

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:

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

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

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

$15\notin\left[\,41,32\ ;\ 109,68\,\right]\ \rightarrow$ 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.

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:

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

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

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.

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.

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

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

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

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. ¿Se puede 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.

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

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

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:

$$I = \left[\begin{pmatrix} \text{proporción} \\ \text{muestral} \end{pmatrix} \pm \begin{pmatrix} \text{error} \\ \text{estimación} \end{pmatrix} \right] \equiv I_{1-\alpha}(p) = \left[\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} \right] \quad \rightarrow \quad \in^2 = \left(z_{\alpha/2} \sqrt{\frac{\hat{p}.\hat{q}}{n}} \right)^2$$

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

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

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

c) Se trata de 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.**

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

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

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

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%.

7. Una central de transformación de productos lácteos recibe diariamente leche de dos granjas A y B. Para analizar la calidad de la leche, que sigue una ley normal, se extraen dos muestras al azar y se analiza el contenido en materia grasa, obteniendo los resultados adjuntos en tantos por ciento:

Granja A	$\bar{x}_A = 8,7\%$	$s_A^2 = 1,02(\%)^2$	$n_A = 33$
Granja B	$\bar{x}_B = 10,9\%$	$s_B^2 = 1,73(\%)^2$	$n_B = 27$

Construir un intervalo de confianza del 95 por 100 para la diferencia del contenido medio en grasa de la leche de ambas granjas.

Solución:

Sea la variable aleatoria $X_A \equiv$ ' Contenido en grasa de la leche de la granja A' que sigue una distribución normal $N(\mu_A, \sigma_A)$. Análogamente, la variable aleatoria $X_B \equiv$ ' Contenido en grasa de la leche de la granja B' que sigue una distribución normal $N(\mu_B, \sigma_B)$.

Se trata de elaborar un intervalo de confianza para la diferencia de medias poblacionales $(\mu_A - \mu_B)$ con varianzas desconocidas y muestras grandes $n_A + n_B = 33 + 27 = 60 > 30$

$$I_{1-\alpha}(\mu_A - \mu_B) = \left[(\bar{x} - \bar{y}) \pm z_{\alpha/2} \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}} \right]$$

$1 - \alpha = 0,95 \quad \alpha = 0,05 \quad \alpha/2 = 0,025 \quad z_{0,025} = 1,96$

$$I_{0,95}(\mu_A - \mu_B) = \left[(8,7 - 10,9) \pm 1,96 \sqrt{\frac{1,02}{33} + \frac{1,73}{27}} \right] = \left[-2,804\%, -1,596\% \right]$$

El intervalo no cubre el 0 % indicando que existe diferencia significativa entre el contenido en grasa de la leche de ambas granjas. Por otra parte, se observa un mayor contenido en grasa en la leche de la granja B.

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 $X_1 \equiv$ 'Producción de maíz del híbrido I' que sigue una distribución normal $N(\mu_1, \sigma_1)$. Análogamente, la variable aleatoria $X_2 \equiv$ ' 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 se puede 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.

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

En el caso,

$$\begin{array}{llll} n_1=5 & \overline{x}_1=85,20 & s_1^2=57,7 & \alpha/2=0,05 \\ n_2=5 & \overline{x}_2=88,6 & s_2^2=9,8 & s_1^2/s_2^2=57,7/9,8=5,89 \end{array}$$

$$F_{0,05\,;\,4\,,\,4}=6,3883 \quad F_{0,95\,;\,4\,,\,4}=1/F_{0,05\,;\,4\,,\,4}=1/6,3883=0,1565$$

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

El intervalo cubre el uno, y concluimos que las varianzas poblacionales son desconocidas e iguales, con una fiabilidad del 90%.

b) Se trata de un intervalo de confianza para la diferencia de medias poblacionales $(\mu_1-\mu_2)$ con varianzas poblacionales desconocidas pero iguales, con muestras pequeñas $n_1+n_2=10<30$.

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

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

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

$$n_1=n_2=5 \quad \overline{x}_1=85,20 \quad \overline{x}_2=88,6 \quad t_{\alpha/2,\,(n_1+n_2-2)}=t_{0,05\,,\,8}=1,860$$

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

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

9. 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:

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

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

Solución:

Sea la variable aleatoria $X_1 \equiv$ 'Vida media del primer esquema' que sigue una distribución normal $N(\mu_1, \sigma_1)$. Análogamente, la variable aleatoria $X_2 \equiv$ '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 se desconoce 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, se construye primero un intervalo de confianza para el cociente de varianzas $(\sigma_1^2 / \sigma_2^2)$, de modo que si el intervalo cubre al punto 1 se puede partir de que las varianzas son desconocidas pero iguales.

Para construir un intervalo de confianza para el cociente de varianzas se emplea la fórmula:

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

$$\begin{cases} s_1^2 = 30^2 = 900 \quad n_1 = 16 \qquad s_2^2 = 17^2 = 289 \quad n_2 = 13 \quad s_1^2 / s_2^2 = 900/289 = 3,114 \\ 1-\alpha = 0,90 \qquad \alpha/2 = 0,05 \end{cases}$$

$$F_{0,05\,;\,15\,,\,12} = 2,6169 \qquad F_{0,95\,;\,15\,,\,12} = 1/F_{0,05\,;\,12\,,\,15} = 1/2,4753 = 0,404$$

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

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

La situación es 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 $n_1 + n_2 < 30$

$$I_{1-\alpha}(\mu_1 - \mu_2) = \left[(\bar{x} - \bar{y}) \pm t_{\alpha/2,\,f} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right]$$

$t_{\alpha/2,f}$ donde f es la aproximación de Welch: $f = \dfrac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\dfrac{(s_1^2/n_1)^2}{n_1+1} + \dfrac{(s_2^2/n_2)^2}{n_2+1}} - 2$

siendo,

$s_1^2 = 30^2 = 900 \quad n_1 = 16 \quad s_1^2/n_1 = 900/16 = 56,25 \quad \left(s_1^2/n_1\right)^2 = 3164,06$

$s_2^2 = 17^2 = 289 \quad n_2 = 13 \quad s_2^2/n_2 = 289/13 = 22,23 \quad \left(s_2^2/n_2\right)^2 = 494,17$

$\dfrac{\left(s_1^2/n_1\right)^2}{17} = 186,12 \qquad \dfrac{\left(s_2^2/n_2\right)^2}{14} = 35,298 \qquad \left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2 = 6159,11$

$$f = \dfrac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\dfrac{(s_1^2/n_1)^2}{n_1+1} + \dfrac{(s_2^2/n_2)^2}{n_2+1}} - 2 = \dfrac{6159,11}{186,12 + 35,298} - 2 = 25,82 \simeq 26$$

$\overline{x}_1 = 1400$ horas $\quad \overline{x}_2 = 1500$ horas $\quad t_{\alpha/2,f} = t_{0,05,26} = 1,706$

$$I_{0,90}(\mu_1 - \mu_2) = \left[(1400 - 1500) \pm 1,706\sqrt{\dfrac{900}{16} + \dfrac{289}{13}}\right] = \left[-115,113\,,\ -84,887\right]$$

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%.

10. 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:

$$I = \left[\overline{d} \pm t_{\alpha/2,(n-1)}\dfrac{s_d}{\sqrt{n}}\right]$$

$d_i = x_i - y_i \qquad \overline{d} = \dfrac{1}{n}\sum_{i=1}^{n} d_i \qquad s_d^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(d_i - \overline{d})^2$

donde $\overline{d}$ 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**

$$n = 10 \quad \bar{d} = 7,40 \quad s_d^2 = 112,1481 \quad s_d = 10,59 \quad t_{\alpha/2,\,(n-1)} = t_{0,025\,,\,9} = 2,262$$

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

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%.