Hoja nº ....

UNIVERSIDAD AUTONOMA DE MADRID

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

ESTADÍSTICA TEÓRICA GRADO GESTIÓN AERONÁUTICA
CONTROL 14 DE DICIEMBRE DE 2015 - 5 ejercicios -

1.- 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?

2.- Se ha realizado una muestra aleatoria simple (m.a.s) de tamaño 17 de una población considerada normal, llegando a la conclusión que la varianza muestal es 4. Calcular la probabilidad $P\left[|\bar{x}-\mu| \leq 0,4325\right]$

3.- Sea una población con media μ de la que se extraen m.a.s. de tamaño n. Considere los siguientes estimadores de la media:

$$\hat{\mu}_1 = \bar{x} \qquad \hat{\mu}_2 = \frac{1}{n+1}\sum_{i=1}^{n} x_i$$

a) Estudiar la insesgadez, la eficiencia relativa y la consistencia de ambos estimadores

b) Elegir el mejor estimador, justificando la respuesta.

UAM UNIVERSIDAD AUTONOMA DE MADRID

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

ESTADÍSTICA TEÓRICA GRADO GESTIÓN AERONÁUTICA
CONTROL 14 DE DICIEMBRE DE 2015 - 5 ejercicios -

4.- Una empresa multinacional desea conocer si con el 95% de fiabilidad existen diferencias significativas entre sus trabajadores en distintos países en el grado de satisfacción en el trabajo. Para ello se toman muestran aleatorias simples de trabajadores, obteniendo los siguientes resultados:

	Satisfacción en el trabajo			
	Muy satisfecho	Satisfecho	Insatisfecho	Muy insatisfecho
A	200 **(242,86)**	300 **(285,71)**	300	100 **(114,29)**
B	300	400	350 **(342,86)**	150 **(152,38)**
C	350 **(283,33)**	300 **(333,33)**	250	150

¿Puede admitirse con un nivel de significación del 5% que la satisfacción en el trabajo es similar en los tres países?

5. Se ha medido el perímetro craneal a niños de edad comprendida entre los dos y tres años, obteniéndose los siguientes datos en centímetros:

38,5	39,2	43,5	41,6	40,3	47,1	39,6	42,5	39,5
43,2	41	42,4	47,5	42,8	38,6	39,4	47,9	39,7
40,4	43	38,8	39,4	40,1	46,2	40,5	45,6	45
38,7	42,7	44,5	45,2	39	45,1	44,2	40,2	

Comprobar si las observaciones craneales son o no distribuidas según una ley normal con un nivel de confianza del 99%.

ESTADÍSTICA TEÓRICA GRADO GESTIÓN AERONÁUTICA
CONTROL 14 DE DICIEMBRE DE 2015 - 5 ejercicios -

1.- 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 $n \leq 30$

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

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

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

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

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.

UNIVERSIDAD AUTONOMA DE MADRID

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

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

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

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

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

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

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

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

2.- Se ha realizado una muestra aleatoria simple (m.a.s) de tamaño 17 de una población considerada normal, llegando a la conclusión que la varianza muestal es 4. Calcular la probabilidad $P\left[|\bar{x}-\mu|\le 0,4325\right]$

Solución:

$X\sim N(\mu,\sigma)$

$n=17<30$

$\sigma_x=2$

$$n\sigma_x^2=(n-1)s_x^2\;\rightarrow\;\frac{\sigma_x}{\sqrt{n-1}}=\frac{s_x}{\sqrt{n}}$$

$$t_{n-1}=\frac{\bar{x}-\mu}{\sigma_x/\sqrt{n-1}}$$

$$t_{17-1}=\frac{\bar{x}-\mu}{2/\sqrt{17-1}}=\frac{\bar{x}-\mu}{2/4}$$

$$P\left[|\bar{x}-\mu|\le 1,22\right]=P\left[\frac{|\bar{x}-\mu|}{2/4}\le\frac{0,4325}{2/4}\right]=$$

$$=P\left[|t_{16}|\le 0,865\right]=P\left[-0,865\le t_{16}\le 0,865\right]=$$

$$=P[t_{16}\ge -0,865]-P[t_{16}\ge 0,865]=P[t_{16}\le 0,865]-P[t_{16}\ge 0,865]=1-2P[t_{16}\ge 0,865]=0,6$$

$n=17<30$

$$t_{17-1}=\frac{\bar{x}-\mu}{2/\sqrt{17-1}}=\frac{\bar{x}-\mu}{2/4}$$

3.- Sea una población con media μ de la que se extraen m.a.s. de tamaño n. Considere los siguientes estimadores de la media:

$$\hat{\mu}_1 = \overline{x} \qquad \hat{\mu}_2 = \frac{1}{n+1}\sum_{i=1}^{n} x_i$$

a) Estudiar la insesgadez, la eficiencia relativa y la consistencia de ambos estimadores

b) Elegir el mejor estimador, justificando la respuesta.

Solución:

a) Insesgadez

Un estimador $\hat{\theta}$ es insesgado (o centrado) cuando se verifica $E(\hat{\theta}) = \theta$

Un estimador $\hat{\theta}$ es sesgado cuando $E(\hat{\theta}) = \theta + \underbrace{b(\hat{\theta})}_{\text{sesgo}} \Rightarrow \underbrace{b(\hat{\theta})}_{\text{sesgo}} = E(\hat{\theta}) - \theta$

Un estimador $\hat{\theta}$ es *asintóticamente insesgado* si su posible sesgo tiende a cero al aumentar el tamaño muestral que se calcula: $\lim\limits_{n\to\infty} b(\hat{\theta}) = 0$

$$E(\hat{\mu}_1) = E(\overline{x}) = E\Big(\frac{1}{n}\sum_{i=1}^{n} x_i\Big) = \frac{1}{n}E\Big(\sum_{i=1}^{n} x_i\Big) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}(n\mu) = \mu$$

$$b(\hat{\mu}_1) = E(\hat{\mu}_1) - \mu = \mu - \mu = 0$$

$$E(\hat{\mu}_2) = E\Big(\frac{1}{n+1}\sum_{i=1}^{n} x_i\Big) = \frac{1}{n+1}E\Big(\sum_{i=1}^{n} x_i\Big) = \frac{1}{n+1}\sum_{i=1}^{n} E(x_i) = \frac{1}{n+1}(n\mu) = \frac{n\mu}{n+1}$$

$$b(\hat{\mu}_2) = E(\hat{\mu}_2) - \mu = \frac{n\mu}{n+1} - \mu = \frac{n\mu - n\mu - \mu}{n+1} = \underbrace{\overbrace{-\frac{\mu}{n+1}}^{\to 0 \text{ cuando "n" aumenta}}}_{\substack{\text{sesgado}\\ \text{asintóticamente}}}$$

- Eficiencia

Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados de un parámetro desconocido θ.

$\hat{\theta}_1$ es *más eficiente* que $\hat{\theta}_2$ si se verifica que $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$

La *eficiencia relativa* se mide por el ratio: $\dfrac{Var(\hat{\theta}_1)}{Var(\hat{\theta}_2)}$

$$V(\hat{\mu}_1) = V(\overline{x}) = V\Big(\frac{1}{n}\sum_{i=1}^{n} x_i\Big) = \frac{1}{n^2}\sum_{i=1}^{n} V(x_i) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

$$V(\hat{\mu}_2) = V(\frac{1}{n+1}\sum_{i=1}^{n} x_i) = \frac{1}{(n+1)^2}\sum_{i=1}^{n} V(x_i) = \frac{1}{(n+1)^2}(n\sigma^2) = \frac{n}{(n+1)^2}\sigma^2$$

$$\text{Eficiencia relativa} \equiv \frac{Var(\hat{\mu}_1)}{Var(\hat{\mu}_2)} = \frac{\sigma^2/n}{n\sigma^2/(n+1)^2} = \frac{(n+1)^2}{n^2} > 1 \mapsto Var(\hat{\mu}_1) > Var(\hat{\mu}_2)$$

El estimador $\hat{\mu}_2$ tiene menor varianza, por lo que es *más eficiente* que $\hat{\mu}_1$

- Consistencia

Un estimador $\hat{\theta}$ *consistente* es un estimador asintóticamente insesgado cuya varianza tiende a cero al aumentar el tamaño muestral.

El estimador $\hat{\theta}$ es *consistente* cuando $\lim_{n\to\infty} E(\hat{\theta}) = \theta$ y $\lim_{n\to\infty} V(\hat{\theta}) = 0$

$$\hat{\mu}_1 \equiv \begin{cases} \lim_{n\to\infty} E(\hat{\mu}_1) = \lim_{n\to\infty} E(\overline{x}) = \mu \\ \lim_{n\to\infty} V(\hat{\mu}_1) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0 \end{cases} \quad \textit{es consistente}$$

$$\hat{\mu}_2 \equiv \begin{cases} \lim_{n\to\infty} E(\hat{\mu}_2) = \lim_{n\to\infty} \left(\mu - \frac{1}{n+1}\mu\right) = \mu \\ \lim_{n\to\infty} V(\hat{\mu}_2) = \lim_{n\to\infty} \left[\frac{n}{(n+1)^2}\sigma^2\right] = 0 \end{cases} \quad \textit{es consistente}$$

b) Elegir uno de los dos en término del error cuadrático medio.

El Error Cuadrático Medio (ECM) de un estimador $\hat{\theta}$ viene definido:

$$ECM(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = V(\hat{\theta}) + \left[\underbrace{E(\hat{\theta}) - \theta}_{\text{sesgo}}\right]^2 \quad \text{sesgo } b(\hat{\theta}) = E(\hat{\theta}) - \theta$$

Si $\underbrace{E(\hat{\theta}) = \theta}_{\text{insesgado}} \Rightarrow ECM(\hat{\theta}) = V(\hat{\theta})$

$$ECM(\hat{\mu}_1) = V(\hat{\mu}_1) + \left[b(\hat{\mu}_1)\right]^2 = \frac{\sigma^2}{n} + 0 = \frac{\sigma^2}{n}$$

$$ECM(\hat{\mu}_2) = V(\hat{\mu}_2) + \left[b(\hat{\mu}_2)\right]^2 = \frac{n}{(n+1)^2}\sigma^2 + \left(\frac{1}{n+1}\mu\right)^2 = \frac{n\sigma^2 + \mu^2}{(n+1)^2}$$

El estimador $\hat{\mu}_1$ será el que presenta menor ECM cuando $ECM(\hat{\mu}_1) \le ECM(\hat{\mu}_2)$

$$V(\hat{\mu}_2) = V(\frac{1}{n+1}\sum_{i=1}^{n} x_i) = \frac{1}{(n+1)^2}\sum_{i=1}^{n} V(x_i) = \frac{1}{(n+1)^2}(n\sigma^2) = \frac{n}{(n+1)^2}\sigma^2$$

$$\text{Eficiencia relativa} \equiv \frac{Var(\hat{\mu}_1)}{Var(\hat{\mu}_2)} = \frac{\sigma^2/n}{n\sigma^2/(n+1)^2} = \frac{(n+1)^2}{n^2} > 1 \mapsto Var(\hat{\mu}_1) > Var(\hat{\mu}_2)$$

El estimador $\hat{\mu}_2$ tiene menor varianza, por lo que es *más eficiente* que $\hat{\mu}_1$

- Consistencia

Un estimador $\hat{\theta}$ *consistente* es un estimador asintóticamente insesgado cuya varianza tiende a cero al aumentar el tamaño muestral.

El estimador $\hat{\theta}$ es *consistente* cuando $\lim_{n\to\infty} E(\hat{\theta}) = \theta$ y $\lim_{n\to\infty} V(\hat{\theta}) = 0$

$$\hat{\mu}_1 \equiv \begin{cases} \lim_{n\to\infty} E(\hat{\mu}_1) = \lim_{n\to\infty} E(\overline{x}) = \mu \\ \lim_{n\to\infty} V(\hat{\mu}_1) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0 \end{cases} \qquad \textit{es consistente}$$

$$\hat{\mu}_2 \equiv \begin{cases} \lim_{n\to\infty} E(\hat{\mu}_2) = \lim_{n\to\infty} \left(\mu - \frac{1}{n+1}\mu\right) = \mu \\ \lim_{n\to\infty} V(\hat{\mu}_2) = \lim_{n\to\infty} \left[\frac{n}{(n+1)^2}\sigma^2\right] = 0 \end{cases} \qquad \textit{es consistente}$$

b) Elegir uno de los dos en término del error cuadrático medio.

El Error Cuadrático Medio (ECM) de un estimador $\hat{\theta}$ viene definido:

$$ECM(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = V(\hat{\theta}) + \left[\underbrace{E(\hat{\theta}) - \theta}_{\text{sesgo}}\right]^2 \qquad \text{sesgo } b(\hat{\theta}) = E(\hat{\theta}) - \theta$$

Si $\underbrace{E(\hat{\theta}) = \theta}_{\text{insesgado}} \Rightarrow ECM(\hat{\theta}) = V(\hat{\theta})$

$$ECM(\hat{\mu}_1) = V(\hat{\mu}_1) + \left[b(\hat{\mu}_1)\right]^2 = \frac{\sigma^2}{n} + 0 = \frac{\sigma^2}{n}$$

$$ECM(\hat{\mu}_2) = V(\hat{\mu}_2) + \left[b(\hat{\mu}_2)\right]^2 = \frac{n}{(n+1)^2}\sigma^2 + \left(\frac{1}{n+1}\mu\right)^2 = \frac{n\sigma^2 + \mu^2}{(n+1)^2}$$

El estimador $\hat{\mu}_1$ será el que presenta menor ECM cuando $ECM(\hat{\mu}_1) \le ECM(\hat{\mu}_2)$

En esta línea,

$$\frac{\sigma^2}{n} \leq \frac{n\sigma^2+\mu^2}{(n+1)^2} = \frac{n\sigma^2}{(n+1)^2} + \frac{\mu^2}{(n+1)^2} \Rightarrow \frac{\sigma^2}{n} - \frac{n\sigma^2}{(n+1)^2} \leq \frac{\mu^2}{(n+1)^2} \Rightarrow$$

$$\Rightarrow \frac{(n+1)^2\sigma^2 - n^2\sigma^2}{n(n+1)^2} \leq \frac{\mu^2}{(n+1)^2} \Rightarrow \frac{(n+1)^2-n^2}{n}\sigma^2 \leq \mu^2 \Rightarrow$$

$$\Rightarrow \frac{2n+1}{n}\sigma^2 \leq \mu^2 \Rightarrow \frac{2n+1}{n} \leq \frac{\mu^2}{\sigma^2}$$

$$\begin{cases} \text{Si } \dfrac{2n+1}{n} \leq \dfrac{\mu^2}{\sigma^2} \mapsto \hat{\mu}_1 \text{ se elige antes que } \hat{\mu}_2 \\[2ex] \text{Si } \dfrac{2n+1}{n} \geq \dfrac{\mu^2}{\sigma^2} \mapsto \hat{\mu}_2 \text{ se elige antes que } \hat{\mu}_1 \end{cases}$$

4.- Una empresa multinacional desea conocer si con el 95% de fiabilidad existen diferencias significativas entre sus trabajadores en distintos países en el grado de satisfacción en el trabajo. Para ello se toman muestran aleatorias simples de trabajadores, obteniendo los siguientes resultados:

	Satisfacción en el trabajo			
	Muy satisfecho	Satisfecho	Insatisfecho	Muy insatisfecho
A	200 **(242,86)**	300 **(285,71)**	300	100 **(114,29)**
B	300	400	350 **(342,86)**	150 **(152,38)**
C	350 **(283,33)**	300 **(333,33)**	250	150

¿Puede admitirse con un nivel de significación del 5% que la satisfacción en el trabajo es similar en los tres países?

Solución:

La hipótesis nula H_0: 'Las proporciones de los trabajadores con los distintos grados de satisfacción son iguales en los tres paises'

Se acepta H_0: $\chi^2_{(k-1).(m-1)} = \sum_{i=1}^{k}\sum_{j=1}^{m}\frac{(n_{ij}-e_{ij})^2}{e_{ij}} = \sum_{i=1}^{k}\sum_{j=1}^{m}\frac{n_{ij}^2}{e_{ij}} - n < \chi^2_{\alpha;\,(k-1).(m-1)}$

Región de rechazo de la hipótesis nula: $R_{rechazo} = \left\{ \chi^2_{(k-1).(m-1)} \geq \chi^2_{\alpha;\ (k-1).(m-1)} \right\}$

Se forma la tabla de contingencia 3 x 4 donde cada frecuencia observada

$(n_{ij})_{i=1,\cdots,k;\ j=1,\cdots,m}$ tiene una frecuencia teórica o esperada $e_{ij} = \dfrac{n_{i\bullet} \times n_{\bullet j}}{n}$

	Satisfacción en el trabajo				Total
	Muy satisfecho	Satisfecho	Insatisfecho	Muy insatisfecho	
A	200 **($e_{11} = 242,86$)**	300 **($e_{12} = 285,71$)**	300 **($e_{13} = 257,14$)**	100 **($e_{14} = 114,29$)**	900 **(900)**
B	300 **($e_{21} = 323,81$)**	400 **($e_{22} = 380,95$)**	350 **($e_{23} = 342,86$)**	150 **($e_{24} = 152,38$)**	1200 **(1200)**
C	350 **($e_{31} = 283,33$)**	300 **($e_{32} = 333,33$)**	250 **($e_{33} = 300$)**	150 **($e_{34} = 133,33$)**	1050 **(1050)**
Total	850	1000	900	400	3150

Estadístico observado: $\chi^2_{(3-1).(4-1)} = \sum_{i=1}^{3}\sum_{j=1}^{4} \dfrac{(n_{ij}-e_{ij})^2}{e_{ij}} = \sum_{i=1}^{3}\sum_{j=1}^{4} \dfrac{n_{ij}^2}{e_{ij}} - n =$

$$= \frac{200^2}{242,86} + \frac{300^2}{285,71} + \frac{300^2}{257,14} + \frac{100^2}{114,29} + \frac{300^2}{323,81} + \frac{400^2}{380,95} + \frac{350^2}{342,86} + \frac{150^2}{152,38} +$$

$$+ \frac{350^2}{283,33} + \frac{300^2}{333,33} + \frac{250^2}{300} + \frac{150^2}{133,33} - 3150 = 49,55$$

Estadístico teórico: $\chi^2_{0,05;\ (3-1).(4-1)} = \chi^2_{0,05;\ 6} = 12,592$

Como $\chi^2_6 = 49,55 > 12,592 = \chi^2_{0,05;\ 6}$ se rechaza la hipótesis nula de homogeneidad de las tres muestras. Es decir, la satisfacción en el trabajo de los empleados de los tres países es significativamente distinta.

5. Se ha medido el perímetro craneal a niños de edad comprendida entre los dos y tres años, obteniéndose los siguientes datos en centímetros:

38,5	39,2	43,5	41,6	40,3	47,1	39,6	42,5	39,5
43,2	41	42,4	47,5	42,8	38,6	39,4	47,9	39,7
40,4	43	38,8	39,4	40,1	46,2	40,5	45,6	45
38,7	42,7	44,5	45,2	39	45,1	44,2	40,2	

Comprobar si las observaciones craneales son o no distribuidas según una ley normal con un nivel de confianza del 99%.

Solución:

Las medidas en centímetros del perímetro craneal ordenadas son:

38,5	39	39,5	40,2	41	42,7	43,5	45,1	47,1
38,6	39,2	39,6	40,3	41,6	42,8	44,2	45,2	47,5
38,7	39,4	39,7	40,4	42,4	43	44,5	45,6	47,9
38,8	39,4	40,1	40,5	42,5	43,2	45	46,2	

El recorrido de la variable $X =$ "centímetros perímetro craneal" es $47{,}9 - 38{,}5 = 9{,}4$.

El número de intervalos es $\sqrt{n} \approx \sqrt{35} \approx 5$, se decide tomar intervalos 5 intervalos (observar que no quede ningún dato fuera de los intervalos).

La amplitud de cada intervalo será aproximadamente $c_i \approx \dfrac{9{,}4}{5} \approx 2$

Se decide hacer 5 intervalos de amplitud 2, empezando en 38 y haciéndolos semiabiertos $[a,b)$. Se forma la tabla:

Intervalos	Marca clase x_i	n_i	$x_i n_i$	$x_i^2 n_i$	p_i	$e_i = np_i$
38 - 40	39	11	429	16731	0,1528	5,348
40 - 42	41	7	287	11767	0,4093	14,3255
42 - 44	43	7	301	12943	0,2653	9,2855
44 - 46	45	6	270	12150	0,1693	5,9255
46 - 48	47	4	188	8836	0,067	2,345
		35	1475	62427		37,2295

Se calcula la media y la desviación típica:

$$\alpha_1 = \mu = \frac{\sum_{i=1}^{5} x_i n_i}{N} = \frac{1475}{35} = 42{,}14 \qquad \alpha_2 = \frac{\sum_{i=1}^{6} x_i^2 n_i}{N} = \frac{62427}{35} = 1783{,}62$$

$\sigma^2 = \alpha_2 - \alpha_1^2 = 1783{,}62 - (42{,}14)^2 = 7{,}84 \qquad \sigma = \sqrt{7{,}84} = 2{,}8$

Considerando una ley normal $N(42{,}14\ ,\ 2{,}8)$ se hallan las probabilidades de cada uno de los intervalos:

$$P[38 \le x < 40) = P\left[\frac{38-42{,}14}{2{,}8} \le \frac{x-42{,}14}{2{,}8} < \frac{40-42{,}14}{2{,}8}\right) = P[-1{,}47 \le z < -0{,}76) =$$
$$= P[0{,}76 \le z < 1{,}47) = 0{,}2236 - 0{,}0708 = 0{,}1528$$

$$P[40 \le x < 42) = P\left[\frac{40-42{,}14}{2{,}8} \le \frac{x-42{,}14}{2{,}8} < \frac{42-42{,}14}{2{,}8}\right) = P[-0{,}76 \le z < -0{,}05) =$$
$$= P[0{,}05 \le z < 0{,}76) = 0{,}4801 - 0{,}0708 = 0{,}4093$$

$$P[42 \le x < 44) = P\left[\frac{42-42{,}14}{2{,}8} \le \frac{x-42{,}14}{2{,}8} < \frac{44-42{,}14}{2{,}8}\right) = P[-0{,}05 \le z < 0{,}66) =$$
$$= P(z \ge -0{,}05) - P(z \ge 0{,}66) = P(z \le 0{,}05) - P(z \ge 0{,}66) = 1 - P(z \ge 0{,}05) - P(z \ge 0{,}66) =$$
$$= 1 - 0{,}4801 - 0{,}2546 = 0{,}2653$$

$$P[44 \le x < 46) = P\left[\frac{44-42{,}14}{2{,}8} \le \frac{x-42{,}14}{2{,}8} < \frac{46-42{,}14}{2{,}8}\right) = P[0{,}66 \le z < 1{,}37) =$$
$$= 0{,}2546 - 0{,}0853 = 0{,}1693$$

$$P[46 \le x < 48) = P\left[\frac{46-42{,}14}{2{,}8} \le \frac{x-42{,}14}{2{,}8} < \frac{48-42{,}14}{2{,}8}\right) = P[1{,}37 \le z < 2{,}09) =$$
$$= 0{,}0853 - 0{,}0183 = 0{,}067$$

Para comprobar si el ajuste se acepta o no, mediante una χ^2, hay que encontrar las frecuencias esperadas $e_i = np_i$. Las frecuencias esperadas de las distintas modalidades no debe ser inferior a cinco.

El intervalo $[46, 48)$ presenta una frecuencia esperada $e_5 = 35 \times 0{,}067 = 2{,}345 < 5$, por lo que hay que agrupar dos intervalos contiguos hasta lograr que la frecuencia esperada sea mayor que cinco.

En este caso se obtiene:

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

Intervalos	n_i	$e_i = np_i$	n_i^2	n_i^2 / e_i
38 - 40	11	5,348	121	22,6253
40 - 42	7	14,3255	49	3,4205
42 - 44	7	9,2855	49	5,2770
44 - 48	10	8,2705	100	12,0912
				43,4140

Se establece la hipótesis nula H_0 : El perímetro craneal sigue una distribución normal

Que se aceptará cuando el estadístico observado $\chi_2^2 = \sum_{i=1}^{5} \frac{n_i^2}{e_i} - n$ es menor que el estadístico teórico o esperado $\chi^2_{\alpha,k-p-1}$

Los grados de libertad es igual al número de clases $(k = 4)$ menos el número de parámetros que ha habido que calcular (μ, σ), $p = 2$, menos 1 al ser excluyentes las modalidades, es decir: $\chi^2_{k-p-1} = \chi^2_{4-2-1} = \chi^2_1$

El estadístico teórico de contraste, con una fiabilidad del 99%, $\alpha = 0{,}01$: $\chi^2_{0{,}01;1} = 6{,}635$

El estadístico observado $\chi_1^2 = \sum_{i=1}^{4} \frac{n_i^2}{e_i} - n = 43{,}4140 - 35 = 8{,}414$

Como se observa que el estadístico observado $\chi_2^2 = 8{,}414$ es mayor que el estadístico teórico $\chi^2_{0{,}01;1} = 6{,}635$, se concluye que las puntuaciones No siguen una distribución $N(42{,}14\ ;\ 2{,}8)$ con una fiabilidad del 99%.