UNIDAD TEMÁTICA 6
ESTIMACIÓN DE PARÁMETROS

ENUNCIADO 1

Las cuotas de ocupación de disco duro (en Mbytes) para diferentes usuarios de una cierta estación de trabajo son:

35 45 47 50 31 30 25 33 35 40 45 47 49 42 40 50 46 55 42 46

*(**1º**) Indica si la muestra es homogénea, en el sentido que no existan puntuaciones atípicas. Siendo dicha muestra representativa de la población de internautas que acceden a los servicios de Internet de dicha estación de trabajo. (**2º**) Calcula el intervalo de confianza del 99 % para la desviación típica que presenta la cuota de ocupación de dicha estación de trabajo. (**3º**) Suponiendo que la distribución de las cuotas de ocupación es normal, ¿cuál es la probabilidad de que la cuota de ocupación de un internauta sea superior a 46 Mbytes o inferior a 34 Mbytes?*

Resolución:

(**2º**) Se trata de un problema de estimación intervalar sobre la desviación típica poblacional (cuya población se considera normal), y en consecuencia hay que utilizar la distribución χ^2 de Pearson, donde el intervalo de confianza viene dado por:

$$\hat{s}\sqrt{\frac{n-1}{\chi_2^2}} \leq \sigma \leq \hat{s}\sqrt{\frac{n-1}{\chi_1^2}}$$

donde $\chi_1^2 = \chi^2_{\alpha=0.5\%,\, \nu=19\ gdl} = 6.8439$ y $\chi_2^2 = \chi^2_{\alpha=99.5\%,\, \nu=19\ gdl} = 38.5824$. Se deduce que el intervalo de confianza buscado es $[5.5409\ Mb,\ 13.1560\ Mb]$.

(**3º**) Dada la naturaleza continua de la variable aleatoria involucrada, X := "cuota de ocupación de disco duro de un usuario", en este caso se ha de proceder en primer lugar a calcular cual es la distribución de probabilidad con la que hay que trabajar. Como se indica en el enunciado se trata de una distribución normal $\mathcal{N}(\mu = 41.65\ Mb, \sigma = 7.8959\ Mb)$ correspondiente a un contexto de estimación intervalar de medias aritméticas siendo el error probable (desviación típica de la estimación) $\sigma_{\hat{\mu}} = \frac{\hat{s}}{\sqrt{n}} = 1.7656\ Mb$. Teniendo en cuenta que $z \triangleq \frac{\overline{x}-\mu}{\sigma_{\hat{\mu}}}$ se tiene que (aprovechando la simetría de la distribución), y dado que se trata de dos sucesos incompatibles (es decir, mutuamente excluyentes):

$$P = \mathbb{P}\left[(X < 34\ Mb) \cup p(X > 46\ Mb)\right] = \mathbb{P}(X < 34\ Mb) + \mathbb{P}(X > 46\ Mb)$$

$$P_1 = \mathbb{P}(X > 46\ Mb) = 1 - \mathbb{P}\left(Z \geq z_1 = \frac{46 - 41.65}{7.8959} \right) = 1 - \mathbb{P}\left(Z \leq z_1 = 0.55092104 \right) = 29.08438695\%$$

$$P_2 = \mathbb{P}(X < 34\ Mb) = \mathbb{P}\left(Z \leq z_2 = \frac{34 - 41.65}{7.8959} \right) = 1 - \mathbb{P}\left(Z \leq z_2 = 0.968861144 \right) = 16.63072318\%$$

y la probabilidad solicitada es $P = 45.71511013\%$.

Nota: Si se enfocara el problema como una distribución t-Student, siendo el número de grados de libertad $v = n - 1 = 19$, se deduce que

$$P_1 = \mathbb{P}(X > 46\ Mb) = 1 - \mathbb{P}\left(t \geq t_1 = \frac{46 - 41.65}{7.8959} \right) = 1 - \mathbb{P}\left(t \leq t_1 = 0.55092104 \right) = 29.4054048\%$$

$$P_2 = \mathbb{P}(X < 34\ Mb) = \mathbb{P}\left(t \leq t_2 = \frac{34 - 41.65}{7.8959} \right) = 1 - \mathbb{P}\left(t \leq t_2 = 0.96886114 \right) = 17.2393484\%$$

con lo que la probabilidad total pedida es $P_1 + P_2 = 46.6447532\%$.

ENUNCIADO 2

Se han realizado 20 mediciones de resistencia a la compresión, en kg/cm^2, de una muestra de cilindros de cemento, a los siete días de secado:

349.09	238.45	385.59	330.00	388.63	348.43	339.85	348.20	361.45	357.33
308.88	196.20	318.99	257.63	299.04	321.47	297.10	218.23	286.23	316.69

*(**1º**) Calcula un intervalo de confianza del 98 % para la media aritmética poblacional? (**2º**) ¿Cuál debería ser el tamaño de la muestra si el error probable de la estimación $\sigma_{\hat{\mu}}$ tuviera que ser 1 kg/cm^2? (**3º**) ¿Existe evidencia estadística de que la desviación típica poblacional es superior a 45.34 kg/cm^2 con un nivel de significación del 1%?*

Resolución:

La estadística descriptiva proporciona los siguientes resultados necearios para las inferencias estadísticas posteriores:

$$\sum_{i=1}^{14} x_i = 6267.48; \sum_{i=1}^{16} (x_i - \overline{x})^2 = 52655.3627 \quad \begin{cases} \overline{x} = 313.374\ kg/cm^2 \\ \hat{s} = 52.6435\ kg/cm^2 \end{cases}$$

Los modelos de probabilidad que intervienen son la distribución t-Student para la estimación de la media aritmética ya que n = 16 < 30, y la distribución χ^2 para la estimación de la desviación típica. De esta forma $t_1 = t_{\alpha=98\%, v=19\ gdl} \equiv t_{99\%, 19\ gdl} = \pm 2.54$ y

$\chi_1^2 = \chi^2_{\alpha=99\%,\, \nu=19\ gdl} = 36.19$. Por otra parte el error probable (desviación típica) de la estimación se obtiene de $\sigma_{\hat{\mu}} = \dfrac{\hat{s}}{\sqrt{n}} = \dfrac{52.6435}{\sqrt{20}} = 11.7714\ kg/cm^2$, con $(n-1)\hat{s}^2 = 19 \times 2632.7681 = 52655.3627\ \left(kg/cm^2\right)^2$. Entonces:

(**1º**) De este modo el intervalo de confianza para la media es:

$$\left[l_\mu, L_\mu\right] = \left[\bar{x} - t_1\sigma_{\hat{\mu}}, \bar{x} + t_1\sigma_{\hat{\mu}}\right] = \left[283.4746\ kg/cm^2, 343.2735\ kg/cm^2\right]$$

(**2º**) De la expresión del error probable se deduce que $n \geq \left(\dfrac{\hat{s}}{A}\right)^2$, siendo $A = 1\ kg/cm^2$ la cota de error que se ha de satisfacer. Operando se obtiene que $n \geq 2771.3349$, y siendo $n \in \mathbb{N}$ habría de tomarse $n \geq 2772$.

(**3º**) En este caso se trata de un contraste de hipótesis de desviación típica de una sola poblacion, unilateral (de cola superior), según se deduce del enunciado; es decir:

$$\begin{cases} H_0: & \sigma_0 = 45.34\ kg/cm^2 \\ H_a: & \sigma > 45.34\ kg/cm^2 \end{cases}$$

El estadístico del contraste es $\chi^2 = \dfrac{(n-1)\hat{s}^2}{\sigma_0^2} = \dfrac{19 \times 52.6435^2}{45.34^2} = \dfrac{52655.3627}{2055.7156} = 25.6141$, que comparado con el valor teórico de la distribución $\chi^2 < \chi_1^2$ (dentro de la región de admisibilidad). En consecuencia, a partir de la muestra presentada existe evidencia estadística para no rechazar la hipótesis nula, y por lo tanto no existe evidencia suficiente para decir que la desviación típica de la población es superior a 45.34 kg/cm^2 con un nivel de significación $\alpha = 1\%$.

ENUNCIADO 3

Se ha estudiado la incidencia de una enfermedad en dos países A y B durante la última década. Los resultados del estudio se recogen en la tabla siguiente:

País	A	B
Tamaño de la muestra	10000	20000
Número de afectados	235	310

*(**1º**) Estima el intervalo de confianza del 95 % para la proporción de afectados del país A. (**2º**) ¿Existe evidencia suficiente de que en el país A la proporción de afectados es mayor que en B en un 0.5 % con un nivel de significación del 2 %? (**3º**) Halla el nivel de significación extremo de la prueba realizada en el apartado (**2º**).*

Resolución:

$$\pi_A \triangleq \hat{\pi}_A = p_A = \frac{235}{10000} = 0.0235\,;\; \pi_B \triangleq \hat{\pi}_B = p_B = \frac{310}{20000} = 0.0155$$

$$\sigma_{\hat{\pi}_A} = \sqrt{\frac{p_A q_A}{n_A}} = \sqrt{\frac{0.0235\times 0.9765}{10000}} = \sqrt{\frac{0.02294775}{10000}} = \sqrt{0.000002295} = 0.001514851$$

$$\sigma_{\hat{\pi}_B} = \sqrt{\frac{p_B q_B}{n_B}} = \sqrt{\frac{0.0155\times 0.9855}{20000}} = \sqrt{\frac{0.015275250}{20000}} = \sqrt{0.000000764} = 0.000873492$$

(1º) Como $\begin{matrix} n_A p_A = 235 > 4 \\ n_A q_A = 9765 \gg 4 \end{matrix} \Rightarrow$ distribución normal /

$$\mathcal{N}(\mu_A = n_A p_A = 235, \sigma_A = \sqrt{n_A p_A q_A} = \sqrt{235\times 0.9765} = 15.1485)$$

Para el nivel de confianza que se sugiere $z_\alpha = z_{95\%} = \pm 1.96$. Al tratarse de un problema de proporciones (equivalentemente, de medias aritméticas) los límites del intervalo de confianza se calculan a partir de

$$\begin{cases} l_A = \hat{\pi}_A - z_\alpha \sigma_{\hat{\pi}_A} = 0.0235 - \overbrace{1.96 * 0.001514851}^{=0.002969109} = 0.020530891 \\ \mathcal{L}_A = \hat{\pi}_A + z_\alpha \sigma_{\hat{\pi}_A} = 0.36 + 0.002953959 = 0.026469109 \end{cases}$$

(2º) El contraste se establece en los siguientes términos:

$$\begin{cases} H_0: & \pi_{A,0} - \pi_{B,0} = 0.05\; (\pi_{A,0} = \pi_{B,0} + 0.005) \\ H_a: & \pi_{A,0} - \pi_{B,0} > 0.05\; (\pi_{A,0} > \pi_{B,0} + 0.005) \end{cases}$$

donde se aplica el modelo de probabilidad normal ya que es

$$n_A p_A = 235 > 4,\; n_A q_A = 9765 \gg 4$$
$$n_B p_B = 310 > 4,\; n_B q_B = 19690 \gg 4$$

El estadístico del contraste es

$$z = \frac{(p_A - p_B) - (\pi_{A,0} - \pi_{B,0})}{\sigma_{\widehat{\pi_A - \pi_B}}} = \frac{(p_A - p_B) - (\pi_{A,0} - \pi_{B,0})}{\sqrt{\frac{p_A q_A}{n_A} + \frac{p_B q_B}{n_B}}} = \frac{0.008 - 0.005}{\sqrt{\frac{0.0235\times 0.9765}{10000} + \frac{0.0155\times 0.9855}{20000}}} =$$

$$\frac{0.003}{\sqrt{0.000002295 + 0.000000763}} = \frac{0.003}{\sqrt{0.0000030577625}} = \frac{0.003}{0.001749} = 1.715613202$$

El valor que separa la región crítica de la zona de admisibilidad para un nivel de significación $\alpha = 2\%$ es

$$z_\alpha \equiv z_{98\%} = 2.053748176$$

Como $z < z_\alpha$, de la muestra proporcionada no existe evidencia estadística para rechazar la hipótesis nula; es decir, no existe evidencia suficiente de que en el país A la proporción de afectados es mayor que en B en un 0.5 % con un nivel de significación del 2 %.

(**3º**) En la prueba realizada se ha establecido un contraste de hipotesis unilateral (de cola superior), de modo que el nivel de significación extrema se alcanzará cuando el nivel de significacion corresponda justamente con el estadístico del contraste; a saber:

$$z = 1.715613202$$

lo que corresponde a un nivel de significación $\alpha = 4.311639525$, puesto que se tiene que

$$p(Z \geq z = 1.715613202)\text{=1-p}(Z \leq 1715613202) \underset{\text{interpolando}}{=} 95.68836048$$

ENUNCIADO 4

Se han comparado dos métodos de producción de unas piezas longitudinales estimando la varianza de las longitudes de dichas piezas obtenidas. El resultado ha sido el siguiente

	n	s^2 (mm)
Método A	11	6.57
Método B	21	4.23

(***a***) *¿Hay evidencia suficiente de que el método B produce menor variabilidad en la longitud de las piezas que el método A con nivel de significación α = 0.05?*
(***b***) *Hallar un intervalo de confianza para la desviación típica σ de la longitud de las piezas producidas con el método B con un coeficiente de confianza del 95%.*

Resolución:

(**a**) Es un problema de dos poblaciones donde entra en juego el cociente de varianzas (de desviaciones típicas para ser exactos), y en consecuencia, la distribución F de Snedecor es el modelo de probabilidad teórico que hay que tomar como referencia.

$$\begin{cases} H_0 : \sigma_A = \sigma_B \\ H_a : \sigma_A > \sigma_B \end{cases}$$

lo que habla de llevar a cabo un contraste unilateral (cola superior).

A partir de las muestras suministradas el estadístico del contraste es:

$$F = \frac{s_A^2}{s_B^2}\,\cancel{\frac{\sigma_B^2}{\sigma_A^2}} = \frac{6.57}{4.23} = 1.5532$$

y el valor crítico es

$$F_1 = F_{95\%,10,20} = 2.35$$

Así pues, como $F < F_1$ no existe evidencia estadística para rechazar la hipótesis nula; es decir, no existe evidencia estadística de que el método B produzca menor variabilidad en la longitud de las piezas que el método A con un nivel de significación del $\alpha = 0.05$.

(B) Es un típico problema de estimación intervalar sobre la desviación típica con lo que el modelo de probabilidad que hay que tener en cuenta es la distribución χ^2.

$$\chi_1^2 = \chi^2_{2.5\%,20\,gdl} \le \frac{(n_B-1)s_B^2}{\sigma_B^2} \le \chi_2^2 = \chi^2_{97.5\%,20\,gdl} \Leftrightarrow$$
$$l = \frac{(n_B-1)s_B^2}{\chi^2_{97.5\%,20\,gdl}} \le \sigma_B^2 \le L = \frac{(n_B-1)s_B^2}{\chi^2_{2.5\%,20\,gdl}}$$

En consecuencia,

$$l = s_B\sqrt{\frac{n_B-1}{\chi^2_{97.5\%,20\,gdl}}} = 2.05670\sqrt{\frac{20}{34.17}} = 2.05670\times 0.7651 = 1.5735$$
$$L = s_B\sqrt{\frac{n_B-1}{\chi^2_{2.5\%,20\,gdl}}} = 2.05670\sqrt{\frac{20}{9.59}} = 2.05670\times 1.4441 = 2.9701$$

siendo en intervalo pedido $1.5735 \le \sigma_B \le 2.9701$.