UNIDAD TEMÁTICA 7
CONTRASTE DE HIPÓTESIS

ENUNCIADO 1

Se ha realizado una encuesta en una población mediante una muestra de 200 personas, resultando 72 fumadores. **(a)** *Estima la proporción de fumadores así como la desviación típica de dicha estimación.* **(b)** *Halla un intervalo de confianza para dicha estimación con un coeficiente de confianza del 95%.* **(c)** *En caso de que deseemos aumentar la precisión de la estimación reduciendo al menos a la mitad la longitud del intervalo de confianza, ¿qué tamaño de muestra debemos utilizar?* **(d)** *¿Hay evidencia suficiente para afirmar que al menos 1/3 de la población es fumadora con un nivel de significación α= 0.05?*

Resolución:

Como se observa se trata de una variable aleatoria binomial (se han tomado la equivalencias $\hat{p} \equiv \hat{\pi}$ y $\sigma_{\hat{p}} \equiv \sigma_{\hat{\pi}}$).

(a) $\hat{p} = \dfrac{72}{200} = 0.36$

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.36 \times 0.64}{200}} = 0.03394113$$

(b) Para obtener el intervalo de confianza aplicamos la expresión:

$$[LI, LS] \approx \left[\hat{p} - z_{1-\frac{\alpha}{2}}\sigma_{\hat{p}}, \quad \hat{p} + z_{1-\frac{\alpha}{2}}\sigma_{\hat{p}}\right]$$

ya que $np = 200 \times 0.36 = 72 > 4$ y $np = 200 \times 0.64 = 128 > 4$. Sustituyendo los diversos valores numéricos se deduce el siguiente intervalo de confianza:

$$IC \equiv [LI, LS] \approx [0.29347671, \quad 0.42652329]$$

siendo $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1{,}95996108$

(c) Se está sugiriendo que el nuevo intervalo de confianza sea al menos $[LI, LS] \approx [0.32673836, \quad 0.39326164]$. Por ello, el valor que debe satisfacer la nueva desviación típica es $\hat{p} + z_{1-\frac{\alpha}{2}}\sigma_{\hat{p}} = 0.39326164$. Es decir, numéricamente se tiene que $\sigma_{\hat{p}} = \dfrac{(0.39326164 \text{ - } \hat{p})}{z_{1-\frac{\alpha}{2}}} = 0.01697056$. Por otra parte, es:

$$n = \frac{\hat{p}\hat{q}}{\sigma_{\hat{p}}^2} = \frac{0.36 \times 0.64}{0.01697056} = 800 \quad personas$$

(d) El contraste de hipótesis que se pide está basado en las siguientes hipótesis:

$$\begin{cases} H_0: & p = \frac{1}{3} = 0.3333 \\ H_a: & p \neq \frac{1}{3} \end{cases}$$

Como $p = \frac{1}{3} \in IC$ sí hay evidencia para afirmar que al menos 1/3 de la población analizada es fumadora con un nivel de significación del 5 %.

ENUNCIADO 2

Se han utilizado dos tipos de soluciones, S1 y S2, para pulir lentes intraoculares para uso humano en operaciones de cataratas. Con la solución S1 se han pulido 300 lentes y se ha observado que 253 de ellas no han producido problemas una vez implantadas. Por otra parte, otras 300 lentes se han pulido con la solución S2 y han sido 196 las lentes que no han causado problemas. ¿Dar un nivel de significación α que indique que de las muestras analizadas hay evidencia estadística para concluir que las dos soluciones son distintas?.

Resolución:

Es un problema de diferencia de proporciones de dos muestras grandes con:

$$p_{S1} = \frac{253}{300} = 0.8433; \; q_{S1} = 0.1567; \; p_{S2} = \frac{196}{300} = 0.6533; \; q_{S2} = 0.3467$$

Es un problema de dos colas (bilateral), ya que el contraste de hipótesis se establece en los siguientes términos:

$$\begin{cases} H_0: & \pi_{S1} = \pi_{S2} \\ H_a: & \pi_{S1} \neq \pi_{S2} \end{cases}$$

siendo posible tratarlo como un caso de grandes muestras ya que:

$$n_{S1}p_{S1} = 252.99 > 5 \quad n_{S2}p_{S2} = 195.99 > 5$$
$$n_{S1}q_{S1} = 47.01 > 5 \quad n_{S2}q_{S2} = 104.01 > 5$$

es decir, el estadístico de referencia es la puntuación típica (el modelo probabilístico del contraste es la distribución normal). El estadístico del contraste es:

$$z=\frac{\hat{\pi}_{S1}-\hat{\pi}_{S2}-(\pi_{S1}-\pi_{S2})}{\sigma_{\hat{\pi}_{S1}-\hat{\pi}_{S2}}}=\frac{\hat{\pi}_{S1}-\hat{\pi}_{S2}-(\pi_{S1}-\pi_{S2})}{\sqrt{\frac{p_{S1}(1-p_{S1})}{n_{S1}}+\frac{p_{S2}(1-p_{S2})}{n_{S2}}}}=$$

$$\frac{0.1900}{0.03458}=5.4952$$

$$\alpha=5\%: z_\alpha=z_{97.5\%}=1.96<z$$
$$\alpha=2.5\%: z_\alpha=z_{98.75\%}=2.243<z$$
$$\alpha=1\%: z_\alpha=z_{99.5\%}=2.575<z$$

Como conclusión: existe evidencia estadística para rechazar la hipótesis nula por pequeño que sea el nivel de significación α; es decir, la solución S1 da lugar a un mayor número de lentes no defectuosas.

ENUNCIADO 3

Se desea evaluar el efecto de una dieta baja en grasa y de ejercicio aeróbico en los niveles de colesterol en sangre, para lo cual se han considerado 15 sujetos entre 35 y 50 años. La tabla muestra el colesterol total inicialmente y después de tres meses de aplicación del programa:

Sujeto	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Antes	265	240	258	295	251	245	287	314	260	279	283	240	238	225	247
Después	229	231	227	240	238	241	234	256	247	239	246	218	219	226	233

¿Los datos de la muestra permiten establecer que una dieta baja en grasas acompañada de ejercicio aeróbico disminuye los niveles de colesterol en sangre con un nivel de significación del 5 %?.

Resolución:

Es un problema de contraste unilateral de hipótesis de diferencia de medias con muestras pequeñas y de pares coincidentes (dada la naturaleza de los datos que intervienen):

$$\begin{cases} H_0: \mu_{antes}=\mu_{despues} \\ H_a: \mu_{antes}>\mu_{despues} \end{cases} \Leftrightarrow \begin{cases} H_0: \mu_{antes}-\mu_{despues}=0 \\ H_a: \mu_{antes}-\mu_{despues}>0 \end{cases}$$

Los datos necesarios para efectuar el contraste son:

$$\sum_i x_i=403 \qquad \sum_i x_i^2=15901$$
$$\widehat{\mu_a-\mu_d}=\bar{x}=26.8667 \qquad s_X=18.3915$$
$$\hat{s}_X=19.0371 \qquad n=15$$

siendo el modelo teórico de probabilidad la t de Student

$$t_1 = \begin{cases} t_{\alpha,\nu} = t_{95\%,14\,gdl} = 1.76 & Modelo\ A \\ t_{\alpha,\nu} = t_{97.5\%,14\,gdl} = 2.14 & Modelo\ B \end{cases}$$

El estadístico de contraste se calcula de acuerdo con

$$t = \frac{(\bar{x}_a - \bar{x}_b) - (\mu_a - \mu_b)}{\hat{s}_{\widehat{\mu_a - \mu_b}} / \sqrt{n}} = \frac{26.8667}{19.0371/\sqrt{15}} = \frac{26.8667}{4.9154} = 5.4658$$

Como $t > t_1$ existe evidencia estadística para rechazar la hipótesis nula con el nivel de confianza que se pide

ENUNCIADO 4

En un intento por reducir la variabilidad en la producción de un método A, un fabricante ha introducido un método B (modificación de A). Para comprobar que dicha modificación efectivamente reduce la variabilidad, se han tomado dos muestras de 25 productos cada una, con los siguientes resultados:

	n	s^2
Método A	25	6.57
Método B	25	3.19

¿Hay evidencia estadística para afirmar que el método B produce menor variabilidad que el método A con un nivel de significación del $\alpha = 0.05$?.

Resolución:

El contraste de hipótesis es $\begin{cases} H_0 : \sigma_A^2 = \sigma_B^2 \\ H_a : \sigma_A^2 > \sigma_B^2 \end{cases}$ (es un problema unilateral de varianzas de dos poblaciones), para lo que se utiliza el estadístico F de Snedecor (lo que implica que las dos poblaciones se supongan normales), con estadístioc:

$$F = \frac{s_A^2}{s_B^2} = \frac{6.57}{3.19} = 2.0596$$

con 20 grados de libertad para A y B. De las tablas de la distribución F de Snedecor se deduce $F_{95\%,20,20} = 2.12$. Es decir, F $< F_{95\%,20,20}$. No hay evidencia estadística para rechazar la hipótesis nula propuesta.

ENUNCIADO 5

Con el fin de disminuir la tolerancia en el diámetro de unas bolas para rodamientos se ha introducido una modificación en el método de fabricación. Se desea contrastar la eficacia del nuevo método. Para ello se han tomado muestras de cada método con el siguiente resultado:

Diámetros (mm.)	**Frecuencia observada Método Antiguo**	**Frecuencia observada Método Nuevo**
0.807	0	0
0.808	1	2
0.809	11	13
0.810	13	11
0.811	5	5
0.812	1	0

(a) *Estima un intervalo de confianza para la varianza del diámetro de las bolas con el Método Antiguo con un coeficiente de confianza de 95%.* **(b)** *Los resultados obtenidos en la muestra, ¿proporcionan evidencia suficiente para afirmar con un nivel de significación α = 0.05 que la varianza del diámetro con el método nuevo es inferior a la obtenida con el método antiguo?*

Resolución:

Para este problema (n = 31) se supondrá que la población de la que se ha extraído la muestra es normal. Los valores de los estadísticos de los dos son:

$$\bar{\phi}_{antiguo} = \frac{1}{n}\sum_{i=1}^{6} f_i\phi_i = 0.8098\ mm$$

$$\hat{s}_{antiguo} = \frac{n}{n-1}\sqrt{\frac{1}{n}\sum_{i=1}^{6} f_i\phi_i^2 - \bar{\phi}_{antiguo}^2} = 0.000872520\ mm$$

$$\bar{\phi}_{nuevo} = \frac{1}{n}\sum_{i=1}^{6} f_i\phi_i = 0.8096\ mm$$

$$\hat{s}_{nuevo} = \frac{n}{n-1}\sqrt{\frac{1}{n}\sum_{i=1}^{6} f_i\phi_i^2 - \bar{\phi}_{nuevo}^2} = 0.000843699\ mm$$

(a) Para fijar ideas se va a considerar un problema de dos colas.

$$Intervalo\ de\ confianza = IC \equiv \left[\frac{(n-1)\hat{s}^2_{antiguo}}{\chi^2_{1-\frac{\alpha}{2},\nu}},\ \frac{(n-1)\hat{s}^2_{antiguo}}{\chi^2_{\frac{\alpha}{2},\nu}}\right]$$

El número de grados de libertad a considerar es $\nu = n-1 = 30$. Entonces, se tiene que para

$$95\ \% \Rightarrow \alpha = 5\ \% : \quad \begin{cases} \chi^2_{1-\frac{\alpha}{2},\nu} = \chi^2_{97.5\,\%,\,30} = 46.9792 \\ \chi^2_{\frac{\alpha}{2},\nu} = \chi^2_{2.5\,\%,\,30} = 16.7908 \end{cases}$$

el intervalo de confianza IC para la varianza del diámetro de las bolas con el método antiguo es

$$IC \equiv \left[0.00000049\, mm^2, \quad 0.00000136\, mm^2\right]$$

(b) En este caso se trata de analizar la razón de las varianzas de los dos métodos:

$$F = \frac{Varianza\ de\ muestra\ mayor}{Varianza\ de\ muestra\ menor} = \frac{\hat{s}^2_{antiguo}}{\hat{s}^2_{nuevo}} = \frac{0.000000761}{0.000000712} = 1.0694864$$

$$95\ \% \Rightarrow \alpha = 5\ \%: \qquad F_{95\ \%,\ 30,\ 30} = 1.84$$

con lo que

$$F = 1.06948649 < F_{95\ \%,\ 30,\ 30} = 1.84$$

de donde se deduce que no hay evidencia estadística para rechazar la hipótesis nula (es decir, no hay evidencia suficiente para afirmar con un nivel de significación del 5 % que la varianza del diámetro con el nuevo método es inferior a la obtenida con el método antiguo).