EJERCICIOS U.D.5 INTRODUCCIÓN A LA INFERENCIA ESTADÍSTICA

5.2- INFERENCIA EN POBLACIONES NORMALES

1. En un estudio sobre los precios alcanzados por los billetes reservados a través de Internet en compañías aéreas de bajo coste, se han recogido los importes de billetes de 17 viajeros al azar que realizaron el trayecto Valencia – Londres en los últimos seis meses con la compañía Brian Air. Para la muestra de viajeros de Brian Air el precio medio de los billetes fue de 24 euros y la varianza del precio de 25 euros^2. Se puede asumir que el precio de los billetes sigue una distribución normal.

a) Brian Air ha lanzado una campaña de propaganda en la que se dice que el precio medio de los billetes Valencia-Londres es de solo 20 euros, asumiendo un riesgo de primera especie del 5% y con la única información de la muestra disponible, ¿se podría aceptar que el precio medio de los billetes es realmente de 20 euros?

H_0 m=20

H_1 m≠20

$t_{calculada} = \dfrac{\bar{x} - m_0}{s/\sqrt{N}} = \dfrac{24-20}{5/\sqrt{17}} = 3{,}298$

que es en valor absoluto > $t^{16}_{tablas\ (\alpha=5\%)} = 2.12 \Rightarrow$

Rechazamos la Hipótesis Nula de que la media es igual a 20.
Con un riesgo de la decisión = riesgo de primera especie = 5%

b) En una muestra de 17 viajeros que realizaron el mismo trayecto en la compañía Isillet Air, la varianza vale 20 euros^2. Con un riesgo de primera especie del 10%, ¿se podría aceptar que la varianza del precio de los billetes es igual en ambas compañías?

Para ver si es admisible H_0: $\sigma_1^2=\sigma_2^2$

a. Se va a obtener el intervalo de confianza para σ_1^2/σ_2^2

b. Se aceptar H_0 si el intervalo contiene el valor 1

Fórmula del intervalo:

$$\left[\frac{s_1^2/s_2^2}{f_2}, \frac{s_1^2/s_2^2}{f_1}\right]$$

donde f_1 y f_2 se obtienen de la tabla de la F: $P(f_1 < F_{N1-1,\ N2-1} < f_2)=0{,}90$ (ya que el riesgo de primera especie es del 10%).

$P(F_{N1-1,\ N2-1} > f_2) = P(F_{16,16} > f_2) = 0{,}05 \Rightarrow$ en la tabla $f_2=2{,}33$

$P(F_{16,16}<f_1)=0{,}05 \Rightarrow f_1= 0{,}43$

Por tanto el intervalo resulta $\left[\frac{25/20}{2{,}33}, \frac{25/20}{0{,}43}\right]$= [0,54, 2,91]

Como el valor 1 está en el intervalo, es admisible la hipótesis nula de igualdad de varianzas del precio de billetes en ambas compañías con un riesgo de primera especie del 10%.

2. Para reajustar los servicios que ofrece, la dirección de una biblioteca virtual desea conocer si el tiempo medio (horas por semana) de utilización de ésta es el mismo para las personas que son socias, como para las que no lo son. Para ello se toman dos muestras aleatorias independientes y se obtiene la siguiente información:

Muestra 1 – Socios	Muestra 2 – NO Socios
$N_1 = 13$	$N_2 = 16$
$S_1 = 2$	$S_2 = 3$

Estudia si es admisible la hipótesis de igualdad de varianzas del tiempo de utilización de la biblioteca virtual en las dos poblaciones. Utiliza un riesgo de primera especie α=2%.

SOLUCIÓN:

Para ver si es admisible H_0: $\sigma_1^2=\sigma_2^2$

a. Se va a obtener el intervalo de confianza para σ_1^2/σ_2^2

b. Se aceptar H_0 si el intervalo contiene el valor 1

Fórmula del intervalo: $$\left[\frac{s_1^2/s_2^2}{f_2}, \frac{s_1^2/s_2^2}{f_1}\right]$$

donde f_1 y f_2 se obtienen de la tabla de la F: $P(f_1 < F_{N1-1,\ N2-1} < f_2)=0{,}98$ (ya que el riesgo de primera especie es del 2%).

$P(F_{N1-1,\ N2-1} > f_2)= P(F_{12,15} > f_2) =0{,}01 \Rightarrow$ en la tabla $f_2=3{,}67$

$P(F_{12,15}<f_1)=0{,}01 \Rightarrow f_1 =0{,}25$.

Por tanto el intervalo resulta $\left[\frac{2^2/3^2}{3{,}67}, \frac{2^2/3^2}{0{,}25}\right]$= [0,12, 1,77]

Como el valor 1 está en el intervalo, es admisible la hipótesis nula de igualdad de varianzas del tiempo de utilización de la biblioteca virtual en las dos poblaciones, con un riesgo de primera especie α=2%.

3. En un sistema de reconocimiento óptico de caracteres (OCR) se han evaluado los porcentajes de acierto de un clasificador basado en una red neuronal.
Dicho clasificador se ha evaluado a partir de un corpus de test que está formado por 10 conjuntos de 1.000 caracteres cada uno. En cada prueba se ha obtenido el porcentaje de aciertos que se produjo en el clasificador, obteniéndose una tabla como la que se muestra a continuación:

Prueba	**Porcentaje de aciertos**
test 1	99
test 2	97
test 3	98
test 4	98
test 5	96
test 6	99
test 7	97
test 8	95
test 9	94
test 10	92

A partir de los datos, contesta a las siguientes preguntas:

a) Realiza un test de hipótesis para saber si se podría garantizar una tasa de acierto del 95% con un riesgo de primera especie de 0,05. Justifica la respuesta.

b) Calcula el intervalo de confianza para la tasa de acierto con nivel de confianza del 95%.

NOTA: Puede suponerse que los datos son normales.

SOLUCIÓN:

a) Para contestar a esta pregunta se debe resolver el contraste de hipótesis con hipótesis nula H_0: m=95
De los datos de la tabla se calcula la media muestral $\overline{X}$ =96,5 y la desviación típica muestral s=2,27

La t calculada es $= \dfrac{\overline{X}-95}{s/\sqrt{N}} = \dfrac{96{,}5-95}{2{,}27/\sqrt{10}} = 2{,}086$

La t de tablas con 9 grados de libertad y riesgo de primera especie α=0,05 es 2,262, como la t calculada es en valor absoluto menor que la t de tablas, se acepta la hipótesis nula. Por tanto, se puede garantizar una tasa de acierto del 95% con un riesgo de primera especie del 5%.

b) El intervalo de confianza para la tasa media de acierto se calcula con la formula:

$$[\overline{X} - t_9^{\alpha=0,05}\frac{s}{\sqrt{N}}, \overline{X} + t_9^{\alpha=0,05}\frac{s}{\sqrt{N}}]$$

$$[96{,}5 - 2{,}262\frac{2{,}27}{\sqrt{10}}, 96{,}5 + 2{,}262\frac{2{,}27}{\sqrt{10}}]$$

[94,874, 98,126] que contiene el valor 95% que se ha aceptado en el apartado anterior.

4. Una marca comercial monta cables USB con una longitud media de 3 metros. Una parte del proceso consiste en cortar el cable (adquirido en rollos de 500 m) mediante una máquina preparada para obtener pedazos de 3 m. En un determinado momento se sospecha que la cortadora está funcionando mal, cortando el cable en trozos de longitud diferente a la esperada.

Con el fin de aclarar esta cuestión, y poder tomar decisiones para que el proceso funcione correctamente, se toma una muestra aleatoria de 20 pedazos que salen de la cortadora. La longitud media de esa muestra es $\overline{X}$=2,96 m y la desviación típica s=0,1m. Con esta información determinar si sería necesario ajustar la máquina (utiliza un valor α=0,05).

Para contestar a esta pregunta se debe resolver el contraste de hipótesis con:
hipótesis nula H_0: m=3 $\Rightarrow$ No ajustar la máquina
hipótesis alternativa H_1: m$\neq$3 $\Rightarrow$ Ajustar la máquina

En el enunciado se da la media muestral $\overline{X}$=2,96 y la desviación típica muestral s=0,1

El intervalo de confianza para la media m (nivel de confianza 95%)

$$[\overline{X} - t_{19}^{\alpha=0,05}\frac{s}{\sqrt{N}}, \overline{X} + t_{19}^{\alpha=0,05}\frac{s}{\sqrt{N}}]$$

$$[2{,}96 - 2{,}093\frac{0{,}1}{\sqrt{20}}, 2{,}96 + 2{,}093\frac{0{,}1}{\sqrt{20}}]$$

[2,91, 3,01]

Como m=3 está en el intervalo se acepta la hipótesis nula. Por tanto no hace falta ajustar la máquina (con un riesgo de primera especie del 5%).

5. Con el fin de determinar la calidad de los programas antes de la fase de depuración, se seleccionaron al azar 14 programas realizados por un programador, registrándose el número de errores que había en cada uno de ellos. Los datos se exponen a continuación:

10 12 7 8 6 10 9 9 10 8 9 11 10 9

a) Calcular el intervalo de confianza para la media del número de errores de los programas (α=0.05)
¿Puede afirmarse, con un nivel de confianza del 95%, que el promedio de errores de los programas es 8? Justifica la respuesta.
b) Calcula un intervalo para σ (con α=10%)

SOLUCIÓN:

a) A partir de los N=14 datos de la muestra se calcula la media muestral $\overline{X}$ =9,14 y la desviación típica muestral s=1,56

El intervalo de confianza para la media m (nivel de confianza 95%)

$$[\overline{X} - t_{13}^{\alpha=0,05}\frac{s}{\sqrt{N}}, \overline{X} + t_{13}^{\alpha=0,05}\frac{s}{\sqrt{N}}]$$

$$[9{,}14 - 2{,}16\frac{1{,}56}{\sqrt{14}}, 9{,}14 + 2{,}16\frac{1{,}56}{\sqrt{14}}]$$

[8,24, 10,04]

Como el valor m=8 no está en el intervalo de confianza, no es admisible una media igual a 8.

b)s=1,56 $\Rightarrow$ s^2=3,12
N=14
Valores de la Chi-Cuadrado con 13 grados de libertad
Para $\alpha/2$=0,05 $\Rightarrow g_2$=22,362
1-$\alpha/2$=0,95 $\Rightarrow$ g_1=5,892
Intervalo de confianza $\left[\sqrt{\frac{13x3{,}12}{22{,}362}}, \sqrt{\frac{13x3{,}12}{5{,}892}}\right]$=[1,346, 2,62]