

Ejercicios Portafolio Inferencia en dos Poblaciones Normales.

Intervalos de confianza y contrastes

- Debéis **entregar los ejercicios por grupos**.
- Para hacer los cálculos tomad, **como mínimo 3 decimales o 3 cifras significativas**.
- Podéis usar las herramientas que consideréis necesarias (Statgraphics, calculadora, tablas, etc) pero tenéis que **justificar todo el desarrollo**.
- **Escribid las fórmulas** que usáis.
- **No solo se valorará el resultado, sino también el desarrollo y la notación usada.**
- Podéis **responder usando un editor de textos o hacerlo a mano** con boli y papel y escanearlo o fotografiarlo luego.

Ejercicio 1

Una empresa informática debe decidir la configuración para los discos duros que instala en los equipos informáticos que vende. Los discos pueden ser configurados según dos tipos de distribución de ficheros (A y B). Para decidir la distribución realiza un experimento con 26 discos configurando 13 de ellos con la distribución de ficheros tipo A y los 13 restantes con la distribución de tipo B. Se ha registrado el tiempo medio de acceso (en segundos) de todas las configuraciones. Los resultados medios correspondientes son los siguientes:

Distrib. A	19,4	20,5	21,4	21,9	23,1	19,7	21,3	19,1	20,5	25,2	21,7	23,9	22,8
Distrib. B	21,6	18,3	14,2	23,5	12,7	19,2	17,3	26,4	15,6	14,9	17,9	14,4	26,6

- a) Estudiar si existen diferencias significativas entre los tiempos medios de acceso de ambos tipos de distribución de ficheros, estableciendo un intervalo de confianza para la diferencia entre los tiempos promedios. Utiliza un riesgo de primera especie $\alpha=1\%$.

Antes de calcular el IC, vamos a plantear la hipótesis

$$H_0 \Rightarrow m_A = m_B$$

$$H_1 \Rightarrow m_A \neq m_B$$

Para calcular los intervalos de confianza, para la distribución A y para la distribución B usaremos la siguiente fórmula

$$IC_{m_1 - m_2}^{nc\%} = \left[(\bar{X}_1 - \bar{X}_2) - z^{\alpha/2} \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}, (\bar{X}_1 - \bar{X}_2) + z^{\alpha/2} \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} \right]$$

Para esto obtenemos el valor que toma la distribución t de Student con $\alpha=0.05\%$ y con 12 grados de libertad.

Después de hacer esto sustituimos y aplicamos los valores correspondientes de la media y la desviación, según si es la distribución A y B

De esta manera obtenemos que el intervalo de confianza de la diferencia de medias es (suponiendo varianzas iguales):

$[-0,925694; 6,75646]$

Como podemos ver la diferencia de medias es bastante grande respecto a los valores que tenemos en A y B

Además como $0 \notin IC$ podemos concluir que $m_A \neq m_B$

b) Estudiar si existen diferencias significativas entre las varianzas de los tiempos de acceso de ambas distribuciones, estableciendo un intervalo de confianza para el cociente entre dichas varianzas. Utiliza un nivel de confianza del 99%.

$$H_0 \Rightarrow \sigma_A^2 = \sigma_B^2$$
$$H_1 \Rightarrow \sigma_A^2 \neq \sigma_B^2$$

Para calcular el intervalo de confianza de las diferencias entre las varianzas vamos a usar la siguiente fórmula

$$IC_{\sigma_1^2/\sigma_2^2}^{nc\%} = \left[\frac{S_1^2}{S_2^2 f_2}, \frac{S_1^2}{S_2^2 f_1} \right]$$

Lo primero que hacemos es obtener el valor de f_1 y f_2 , usando la distribución F de Snecdor

Sustituyendo obtenemos que el IC es $[0,0314075; 0,756019]$

Como $1 \notin IC$, podemos concluir que las diferencias son diferentes

NOTA: obtener los dos intervalos de confianza “a mano” y comprobar los resultados con los proporcionados por *Statgraphics centurión*.

Ejercicio 2

En un estudio de opinión realizado en una ciudad, se encuestó a una muestra de 500 personas para determinar su preferencia entre dos marcas de smartphones: *TecnoPhone* y *SmartStar*. Los resultados obtenidos fueron los siguientes:

De las 500 personas encuestadas, 320 manifestaron preferir *TecnoPhone*. De las 500 personas encuestadas, 180 manifestaron preferir *SmartStar*.

Realiza una prueba de hipótesis con un nivel de significación del 1% para determinar si hay suficiente evidencia para afirmar que el porcentaje de personas que prefieren la marca *TecnoPhone* es igual que el porcentaje de personas que prefieren la marca *SmartStar* en la población.

Antes de calcular el intervalo de confianza, planteamos la hipótesis

$$\begin{aligned}H_0 &=> \widehat{P}_1 = \widehat{P}_2 \\H_1 &=> \widehat{P}_1 \neq \widehat{P}_2\end{aligned}$$

Para calcular diferencia de proporciones, lo primero que vamos a hacer es calcular las proporciones muestrales

$$\widehat{P}_1 = 0,64, \text{ las personas que prefieren TecnoPhone}$$

$$\widehat{P}_2 = 0,36, \text{ las personas que prefieren Smartstar}$$

A partir de aquí, aplicamos la siguiente fórmula

$$IC_{\widehat{P}_1 - \widehat{P}_2}^{nc\%} = [(\widehat{P}_1 - \widehat{P}_2) - z^{\alpha/2} S_{\widehat{P}_1 - \widehat{P}_2}, (\widehat{P}_1 - \widehat{P}_2) + z^{\alpha/2} S_{\widehat{P}_1 - \widehat{P}_2}]$$

De esta manera y sustituyendo con los valores adecuados, y previamente habiendo calculado el estadístico Z con $\alpha = 1\%$, obtenemos que el intervalo de confianza es

$$[0,201803; 0,358197]$$

Como $0 \notin IC$, rechazamos H_0

Ejercicio 3

Supongamos la siguiente situación: se desea comparar dos algoritmos de inversión de grandes matrices para ver cuál resulta en promedio más rápido. El tiempo de resolución depende de las características de cada matriz (tamaño, nº de ceros, posición de los mismos...) por lo que fluctúa de un caso a otro. Se invirtieron 12 matrices, cuyos valores se generaron aleatoriamente, mediante los dos algoritmos constatando los siguientes tiempos de resolución en segundos.

Matriz	ALGORITMO 1	ALGORITMO 2
1	2,3	2,1
2	4,1	4,7
3	5,6	5,8
4	3,9	4,7
5	1,2	1,7
6	3,8	3,8
7	6,9	7,7
8	4,4	4,9
9	2,4	3,1
10	3,1	2,8
11	8,6	9,2
12	2,1	2,6

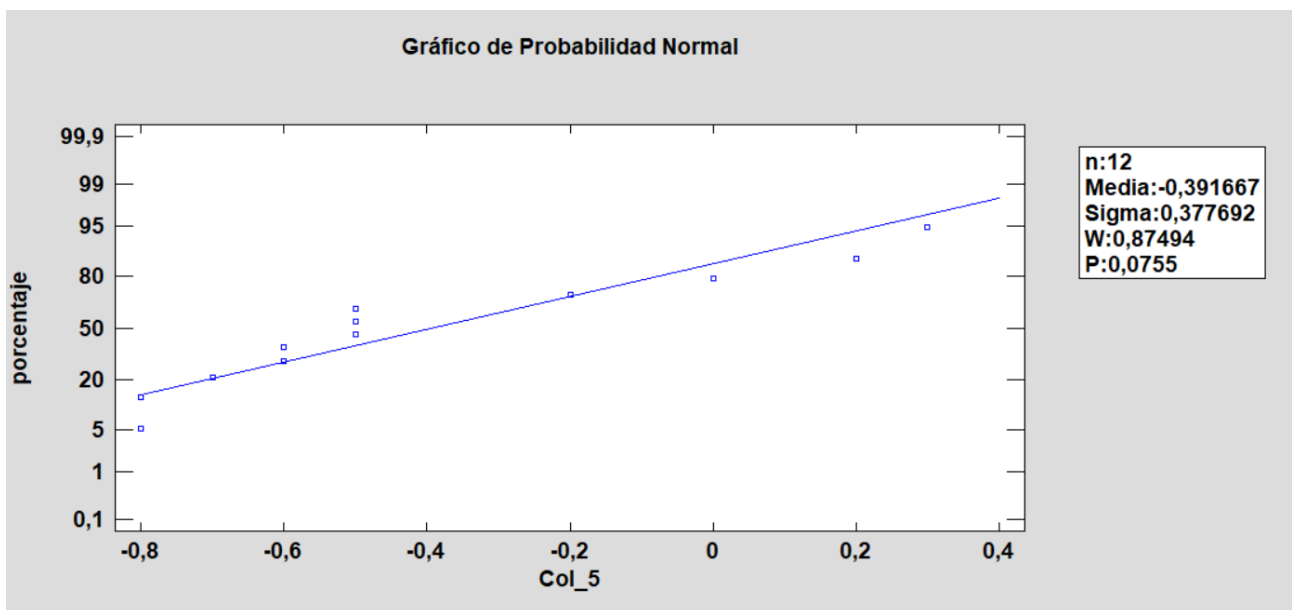
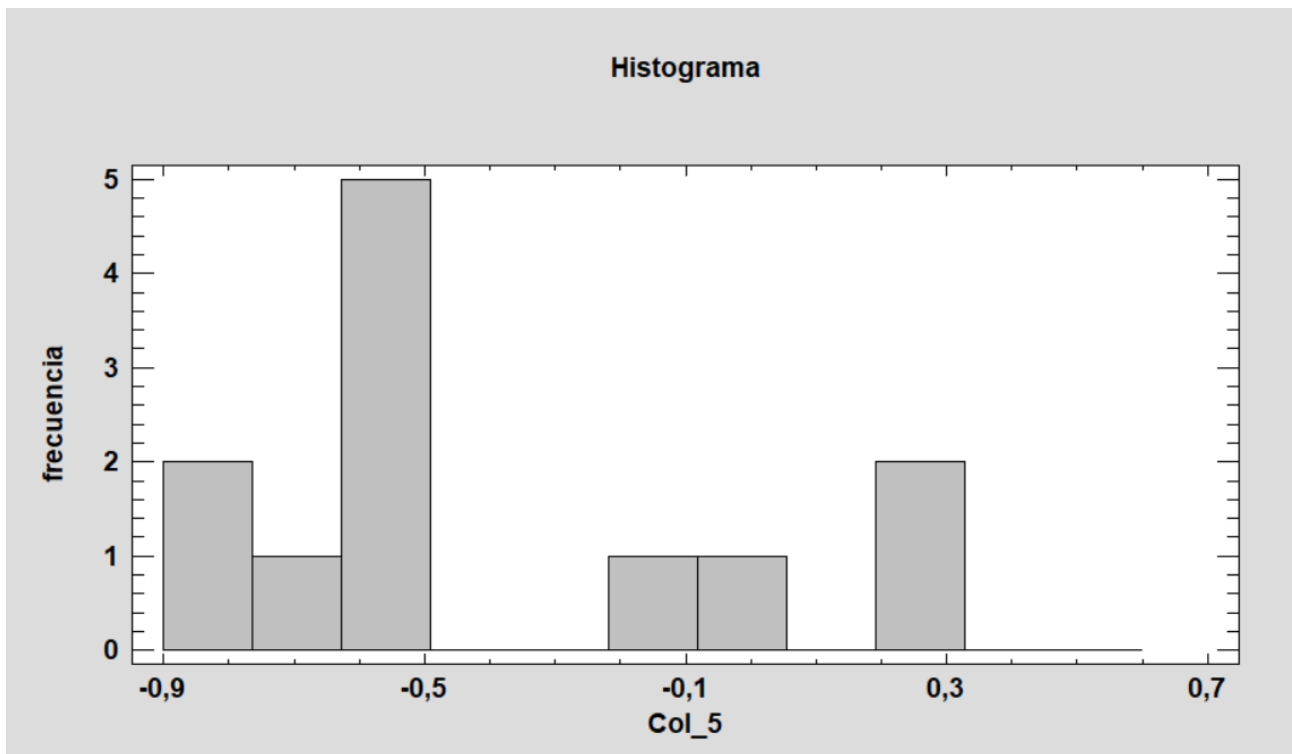
Responde a las siguientes preguntas:

- Justifica cómo se abordaría el análisis a partir del diseño planteado de las pruebas (¿se utilizará el test de comparación de medias o sería más adecuado utilizar datos apareados? ¿sería indiferente?).

Para abordar el análisis del diseño planteado usaremos muestras apareadas, ya que los algoritmos se aplican sobre la misma matriz, de esta manera le daríamos más potencia al experimento

- Estudiar si es razonable suponer que la diferencia de tiempos de resolución entre ambos algoritmos varía normalmente de unas matrices a otras.

Obteniendo la diferencia de tiempo podemos observar que no sigue una distribución normal, como podemos ver en el siguiente histograma y en el gráfico de probabilidad normal



- c) Estudiar si existen diferencias significativas entre las velocidades promedias de ambos algoritmos aplicando el test correspondiente y estableciendo un intervalo de confianza para la diferencia entre los tiempos promedios.

Para calcular el intervalo de confianza de la diferencia de media, primero hallamos la diferencia de medias y después aplicamos la siguiente fórmula con $\alpha = 1\%$ para obtener el IC

$$IC_m \Rightarrow \left[\bar{x} - t_{\frac{\alpha}{2}, N-1} \frac{S}{\sqrt{N}}, \bar{x} + t_{\frac{\alpha}{2}, N-1} \frac{S}{\sqrt{N}} \right]$$

[0,242168; 0,776388]

Como $0 \notin IC$ concluimos que hay una diferencia significativa entre las medias