

▼ Actividad: Diferencia entre medias poblacionales

Andrés Alejandro Guzmán González - A01633819

```
1 import pandas as pd
2 import numpy as np
3 from scipy.stats import norm, t
```

▼ Problema 1:

Un científico de datos está analizando los niveles de sodio en dos lotes diferentes de un mismo producto. El científico quiere determinar si los niveles de sodio son iguales para ambos lotes, por lo tanto, recabó las siguientes dos muestras de datos

Lote A -(número de muestras 15): Nivel de Sodio (mg) -> 180, 160, 170, 190, 200, 175, 185, 195, 180, 170, 190, 185, 200, 175, 165

Lote B -(número de muestras 20): Nivel de Sodio (mg) -> 210, 215, 220, 225, 230, 215, 220, 225, 230, 235, 210, 215, 220, 225, 230, 215, 220, 225, 230, 23

La varianza poblacional para el lote A es 57.05 y para el lote B 34.63. Con un nivel de confianza del 85% determina si ambos lotes tienen en promedio el mismo nivel de sodio.

```
1 dataA = {
2     'Lote_A': [180, 160, 170, 190, 200, 175, 185, 195, 180, 170, 190, 185, 200, 175, 165],
3 }
4 dataB = {
5     'Lote_B': [210, 215, 220, 225, 230, 215, 220, 225, 230, 235, 210, 215, 220, 225, 230, 215, 220, 225, 230, 23]
6 }
7
8 dfA = pd.DataFrame(dataA)
9 dfB = pd.DataFrame(dataB)
```

Prueba Z Hipotesis nula $H_0 : \mu_1 = \mu_2$ Estadístico de prueba

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

El estadístico de prueba tiene una distribución normal estándar.

```
1 prom_A = dfA['Lote_A'].mean()
2 m = 15
3 sigma_A = 57.05
4
5 prom_B = dfB['Lote_B'].mean()
6 n = 20
7 sigma_B = 34.63
8
9 niv_conf = 0.85
10
11 Z = (prom_A - prom_B) / np.sqrt((sigma_A / m) + (sigma_B / n))
12 Z

-12.9925921892129

1 val_critico = norm.ppf(1 - (1 - niv_conf) / 2)
2
3 # Comparación de Z con el valor crítico
4 if Z < -val_critico or Z > val_critico:
5     conclusion = "Rechazamos la hipótesis nula"
6 else:
7     conclusion = "No rechazamos la hipótesis nula"
8 print (' Valor de Z: ',Z,'\n Valor crítico: ',val_critico,'\n',conclusion)

Valor de Z: -12.9925921892129
Valor crítico: 1.4395314709384563
Rechazamos la hipótesis nula
```

▼ Problema 4:

Se han tomado dos muestras del número de días que tardan los egresados de las universidades A y B en encontrar trabajo

Universidad A: -> 180, 200, 190, 210, 175, 185, 195, 180, 205, 190, 200, 185, 210, 175, 195

Universidad B: -> 210, 215, 220, 225, 230, 215, 220, 225, 230, 235, 210, 215, 220, 225, 230, 235

Utilizando un nivel de confianza del 95%, determina si hay evidencia estadística suficiente para concluir que existe una diferencia significativa en el tiempo promedio de búsqueda de empleo entre ambas universidades.

- Utiliza la distribución t-student

Entrega: Entrega tu actividad en canvas mostrando todo el procedimiento necesario para dar respuesta a los ejercicios planteados.

```
1 Uni_A = np.array( [180, 200, 190, 210, 175, 185, 195, 180, 205, 190, 200, 185, 210, 175, 195])
2 Uni_B = np.array( [210, 215, 220, 225, 230, 215, 220, 225, 230, 235, 210, 215, 220, 225, 230, 235])
3
4

1 niv_conf_2 = 0.95
2
3 med_A = np.mean(Uni_A)
4 med_B = np.mean(Uni_B)
5 desv_est_A = np.std(Uni_A, ddof=1)
6 desv_est_B = np.std(Uni_B, ddof=1)
7
8 print('media A: ',med_A,'    media B: ',med_B,'\nDesv. A: ', desv_est_A, '    Desv. B: ', desv_est_B)

media A:  191.66666666666666    media B:  222.5
Desv. A:  11.751393027860065    Desv. B:  8.16496580927726

1 dif_error_est = np.sqrt((desv_est_A**2 / len(Uni_A)) + (desv_est_B**2 / len(Uni_B)))
2 dif_error_est

3.656913435264209

1 t_val = (med_A - med_B) / dif_error_est
2 grad_lib = len(Uni_A) + len(Uni_B)
3 t_critico = t.ppf(1 - (1 - niv_conf_2) / 2, grad_lib)
4
5 if t_val < -t_critico or t_val > t_critico:
6     conclusion = "Rechazamos la hipótesis nula"
7 else:
8     conclusion = "No Rechazamos la hipótesis nula"
9 print (' Valor de t: ',t_val,'\n Valor crítico: ',t_critico,'\n',conclusion)

Valor de t:  -8.431518513947449
Valor crítico:  2.0395134463964077
Rechazamos la hipótesis nula
```