

Actividad: Diferencia entre medias poblacionales

Tania Sayuri Guizado Hernández A01640092

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.

```
In [1]: #Cargamos las librerías necesarias para la actividad
import numpy as np
import pandas as pd
from scipy import stats
from scipy.stats import norm
from statsmodels.stats.weightstats import ztest
```

Prueba Z

Hipotesis nula

$$H_0 : \mu_1 = \mu_2$$

Estadística de prueba

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

```
In [2]: # Datos
lote_a = [180, 160, 170, 190, 200, 175, 185, 195, 180, 170, 190, 185, 200, 175, 165]
lote_b = [210, 215, 220, 225, 230, 215, 220, 225, 230, 235, 210, 215, 220, 225, 230, 215, 220, 225]

# Varianzas poblacionales
var_a = 57.05
var_b = 34.63

# Nivel de confianza
confidence_level = 0.85

# Medias y tamaño del arreglo
mean_a = np.mean(lote_a)
mean_b = np.mean(lote_b)
n_a = len(lote_a)
n_b = len(lote_b)

# Realizar prueba Z
z_statistic, _ = ztest(lote_a, lote_b, value=0, alternative='two-sided')

# Valor crítico z basado en el nivel de confianza
z_critical = stats.norm.ppf(1 - (1 - confidence_level) / 2)

# Comparar el estadístico z con el valor crítico
if abs(z_statistic) > z_critical:
    print('Valor de z_statistic: ', z_statistic)
    print('Valor de z_critical: ', z_critical)
    print("Se rechaza la hipótesis nula. Hay evidencia estadística para afirmar que los niveles de sodio no son iguales en ambos lotes.")
else:
    print("No se rechaza la hipótesis nula. No hay evidencia estadística para afirmar que los niveles de sodio no son iguales en ambos lotes.")
```

Valor de z_statistic: -12.147837775431165

Valor de z_critical: 1.4395314709384563

Se rechaza la hipótesis nula. Hay evidencia estadística para afirmar que los niveles de sodio no son iguales en ambos lotes.

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.

Hint: Utiliza la distribución t-student

```

In [3]: # Datos
universidad_a = np.array([180, 200, 190, 210, 175, 185, 195, 180, 205, 190, 200, 185, 210, 175, :
universidad_b = np.array([210, 215, 220, 225, 230, 215, 220, 225, 230, 235, 210, 215, 220, 225, :

# Nivel de confianza
confidence_level = 0.95

# Medias, desviación y tamaños
mean_diff = np.mean(universidad_a) - np.mean(universidad_b)
n = len(universidad_a)
m = len(universidad_b)
s_a = np.std(universidad_a, ddof=1)
s_b = np.std(universidad_b, ddof=1)
s_diff = np.sqrt((s_a**2/n)+(s_b**2/m))

# Estadístico de prueba t
t = mean_diff / s_diff

# Definir el valor crítico de acuerdo al nivel de confianza y los grados de libertad
alpha = 1 - confidence_level
degrees_freedom = len(universidad_a) + len(universidad_b)

# Valor crítico
t_critical = stats.t.ppf(1 - alpha / 2, degrees_freedom)

# Comparación
if abs(t) > t_critical:
    print('Valor de t: ',t)
    print('Valor de t_critical: ',t_critical)
    print("Se rechaza la hipótesis nula. Hay evidencia estadística para afirmar que existe una d
else:
    print("No se rechaza la hipótesis nula. No hay evidencia estadística para afirmar que hay una

```

Valor de t: -8.431518513947449

Valor de t_critical: 2.0395134463964077

Se rechaza la hipótesis nula. Hay evidencia estadística para afirmar que existe una diferencia significativa en el tiempo promedio de búsqueda de empleo entre ambas universidades.