**SCRIPT DE INGRESO**

|  |
| --- |
| import pandas as pd  import matplotlib.pyplot as plt  import numpy as np  import statistics  import scipy.stats as stats  # Ruta completa del archivo CSV en la unidad C:  archivo\_csv = r'C:\Users\USER\OneDrive\Desktop\PF\ECH\_2022 - BD Proyecto Final PyE 2023.csv'  # Leer el archivo CSV con separador (;)  datos\_csv = pd.read\_csv(archivo\_csv, sep=';')  # Acceder a las columnas y convertirlas en vector  ID=datos\_csv['ID']  AÑO=datos\_csv['anio']  MES=datos\_csv['mes']  SEXO=datos\_csv['Sexo']  EDAD=datos\_csv['Edad']  REGION=datos\_csv['region']  PEA=datos\_csv['PEA']  DESEMPLEO=datos\_csv['Desempleo']  filtro = datos\_csv[(PEA == 1) & (DESEMPLEO == 0)]  SALARIO = filtro['Salario']  # A1.a) Tasa de desempleo para la muestra  td=DESEMPLEO.sum()/PEA.sum()\*100  print("La tasa de desempleo es de {:.2f}%".format(td))  # A1.b) Gráfico de tasa de desempleo por edad  filtro = datos\_csv[(EDAD >= 14) & (EDAD <= 17)]  td1 = filtro['Desempleo'].sum()/filtro['PEA'].sum()\*100  filtro = datos\_csv[(EDAD >= 18) & (EDAD <= 25)]  td2 = filtro['Desempleo'].sum()/filtro['PEA'].sum()\*100  filtro = datos\_csv[(EDAD >= 26) & (EDAD <= 40)]  td3 = filtro['Desempleo'].sum()/filtro['PEA'].sum()\*100  filtro = datos\_csv[(EDAD >= 41)]  td4 = filtro['Desempleo'].sum()/filtro['PEA'].sum()\*100  plt.bar(["14-17", "18-25", "26-40", "Más de 40"], [td1, td2, td3, td4])  plt.title("Tasa de desempleo por rango de edad")  plt.xlabel("Rango de edad")  plt.ylabel("Tasa de desempleo (%)")  plt.show()  # A2.a) Histograma de salarios  plt.hist(SALARIO, bins=100, edgecolor='blue', density=True)  plt.title("Histograma de Salarios")  plt.ylabel("Densidad Relativa")  plt.show()  # A2.b) Elaborar y corregir Boxplot de salarios  plt.boxplot(SALARIO)  plt.title("Boxplot de Salarios")  plt.ylabel("Salario")  plt.show()  Q1 = np.quantile(SALARIO, 0.1)  Q3 = np.quantile(SALARIO, 0.9)  IQR = Q3 - Q1  li = Q1 - 1.5 \* IQR  ls = Q3 + 1.5 \* IQR  SALARIO\_corregido = np.where((SALARIO < li) | (SALARIO > ls), np.nan, SALARIO)  SALARIO\_corregido = SALARIO\_corregido[~np.isnan(SALARIO\_corregido)]  plt.boxplot(SALARIO\_corregido)  plt.title("Boxplot de Salarios Corregidos")  plt.ylabel("Salario")  plt.show()  # A2.c) Calcular media, mediana y moda de salarios.  print("Media de salarios:", np.mean(SALARIO))  print("Mediana de salarios:", np.median(SALARIO))  print("Moda de salarios:", statistics.mode(SALARIO))  # A2.d) Calcular mínimo, máximo y cuartiles de salario.  salario\_minimo = np.min(SALARIO)  salario\_maximo = np.max(SALARIO)  cuartiles = np.percentile(SALARIO, [25, 50, 75])  print("Mínimo salario: ", np.min(SALARIO))  print("Máximo salario: ", np.max(SALARIO))  print("Cuartiles: ", np.percentile(SALARIO, [25, 50, 75]))  # A2.e) Presentar boxplot de salario por género y región  filtro1 = datos\_csv[SEXO == 1]  filtro2 = datos\_csv[SEXO == 2]  plt.figure(figsize=(8, 6))  plt.boxplot([filtro1['Salario'], filtro2['Salario']], labels=['Varones', 'Mujeres'])  plt.title("Boxplot de Salarios por Género")  plt.xlabel("Género")  plt.ylabel("Salario")  plt.show()  filtro1 = datos\_csv[REGION == 1]  filtro2 = datos\_csv[REGION != 1]  plt.figure(figsize=(8, 6))  plt.boxplot([filtro1['Salario'], filtro2['Salario']], labels=['Montevideo', 'Interior'])  plt.title("Boxplot de Salarios por Región")  plt.xlabel("Región")  plt.ylabel("Salario")  plt.show()  # B1) Estimar el desempleo del total de la población  print("Desempleo estimado: ", int(td/100\*1757161))  # B2) Elaborar IC para la variable desempleo al 95%  zo = stats.norm.ppf(1 - 0.05/2)  ET = (td/100\*(1-td/100)/PEA.sum())\*\*0.5  LCi = td/100-ET\*zo  LCs = td/100+ET\*zo  print("IC\_desempleo: ", [int(LCi\*1757161),int(LCs\*1757161)])  # C1) Prueba de Hipótesis - Desempleo  zo = stats.norm.ppf(0.05)  ET = (7/100\*(1-7/100)/PEA.sum())\*\*0.5  RAi = 7/100+ET\*zo  if td >= 7:  print("La tasa de desempleo aumentó respecto del 2021")  else:  print("La tasa de desempleo disminuyó respecto del 2021")  # C2) Prueba de Hipótesis - Salario  filtro1 = datos\_csv[(PEA == 1) & (DESEMPLEO == 0) & (SEXO == 1)]  filtro2 = datos\_csv[(PEA == 1) & (DESEMPLEO == 0) & (SEXO == 2)]  zo = stats.norm.ppf(1 - 0.01/2)  n1 = len(filtro1)  n2 = len(filtro2)  S1 = np.std(filtro1['Salario'])  S2 = np.std(filtro2['Salario'])  ET = (S1\*\*2/n1+S2\*\*2/n2)\*\*0.5  RAi = 0-ET\*zo  RAs = 0+ET\*zo  m1 = np.mean(filtro1['Salario'])  m2 = np.mean(filtro2['Salario'])  if ((m1-m2 >= RAi) & (m1-m2 <= RAs)):  print("Los salario de varones y mujeres son iguales")  else:  print("Los salario de varones y mujeres son distintos") |

**SCRIPT DE SALIDA**

|  |
| --- |
| La tasa de desempleo es: **7.54**%  La media de salarios: **43271.93993710942**  La mediana de salarios: **32830.67**  La moda de salarios: **20000.0**  El salario mínimo es: **0.0**  El salario máximo es: **9765833.0**  Los cuartiles de salario son: [**18483.35** **32830.67** **52659.67**]  El desempleo estimado es: **132564**  El IC para el desempleo: [**127046**, **138082**]  La tasa de desempleo aumentó respecto **del** **2021**  Los salario de varones y mujeres son distintos |

# ESTADÍSTICA DESCRIPTIVA

## DESEMPLEO

**a. Calcular tasa de desempleo para la muestra.**

El script de cálculo para este apartado es

|  |
| --- |
| # A1.a) Tasa de desempleo para la muestra  td=DESEMPLEO.sum()/PEA.sum()\***100**  **print**("La tasa de desempleo es: {:.2f}%".format(td)) |

La salida correspondiente resulta

|  |
| --- |
| La tasa de desempleo es: **7.54**% |

**b. Presentar gráfico que muestre la tasa de desempleo diferenciando por rango de edad (14-17; 18-25; 26-40; más de 40 años).**

El script de cálculo para este apartado es

|  |
| --- |
| # A1.b) Gráfico de tasa de desempleo por edad  filtro = datos\_csv[(EDAD >= **14**) & (EDAD <= **17**)]  td1 = filtro['Desempleo'].sum()/filtro['PEA'].sum()\***100**  filtro = datos\_csv[(EDAD >= **18**) & (EDAD <= **25**)]  td2 = filtro['Desempleo'].sum()/filtro['PEA'].sum()\***100**  filtro = datos\_csv[(EDAD >= **26**) & (EDAD <= **40**)]  td3 = filtro['Desempleo'].sum()/filtro['PEA'].sum()\***100**  filtro = datos\_csv[(EDAD >= **41**)]  td4 = filtro['Desempleo'].sum()/filtro['PEA'].sum()\***100**  plt.bar(["14-17", "18-25", "26-40", "Más de 40"], [td1, td2, td3, td4])  plt.title("Tasa de desempleo por rango de edad")  plt.xlabel("Rango de edad")  plt.ylabel("Tasa de desempleo (%)")  plt.show() |

La salida correspondiente resulta

A picture containing text, screenshot, diagram, plot

Description automatically generated

## SALARIO

**a. Elaborar histograma de salarios.**

El script de cálculo para este apartado es

|  |
| --- |
| # A2.a) Histograma de salarios  plt.hist(SALARIO, bins=**100**, edgecolor='blue', density=True)  plt.title("Histograma de Salarios")  plt.xlabel("Salarios")  plt.ylabel("Frecuencia Relativa")  plt.show() |

La salida correspondiente resulta

A picture containing text, screenshot, diagram, line

Description automatically generated

**b. Elaborar Box-plot para toda la muestra.**

El script de cálculo para este apartado es

|  |
| --- |
| # A2.b) Elaborar y corregir Boxplot de salarios  plt.boxplot(SALARIO)  plt.title("Boxplot de Salarios")  plt.ylabel("Salario")  plt.show()  Q1 = np.quantile(SALARIO, **0.1**)  Q3 = np.quantile(SALARIO, **0.9**)  IQR = Q3 - Q1  li = Q1 - **1.5** \* IQR  ls = Q3 + **1.5** \* IQR  SALARIO\_corregido = np.where((SALARIO < li) | (SALARIO > ls), np.nan, SALARIO)  SALARIO\_corregido = SALARIO\_corregido[~np.isnan(SALARIO\_corregido)]  plt.boxplot(SALARIO\_corregido)  plt.title("Boxplot de Salarios Corregidos")  plt.ylabel("Salario")  plt.show() |

La salida correspondiente resulta

A picture containing text, screenshot, diagram, rectangle

Description automatically generated

A picture containing text, screenshot, diagram, line

Description automatically generated

**c. Calcular media, mediana y moda de salarios.**

El script de cálculo para este apartado es

|  |
| --- |
| # A2.c) Calcular media, mediana y moda de salarios.  **print**("**\n**La media de salarios:", np.mean(SALARIO))  **print**("La mediana de salarios:", np.median(SALARIO))  **print**("La moda de salarios:", statistics.mode(SALARIO))  td=DESEMPLEO.sum()/PEA.sum()\***100**  **print**("La tasa de desempleo es: {:.2f}%".format(td)) |

La salida correspondiente resulta

|  |
| --- |
| La media de salarios: **43271.93993710942**  La mediana de salarios: **32830.67**  La moda de salarios: **20000.0** |

**d. Calcular mínimo, máximo y cuartiles de salario.**

El script de cálculo para este apartado es

|  |
| --- |
| # A2.d) Calcular mínimo, máximo y cuartiles de salario.  salario\_minimo = np.min(SALARIO)  salario\_maximo = np.max(SALARIO)  cuartiles = np.percentile(SALARIO, [**25**, **50**, **75**])  **print**("**\n**El salario mínimo es: ", np.min(SALARIO))  **print**("El salario máximo es: ", np.max(SALARIO))  **print**("Los cuartiles de salario son: ", np.percentile(SALARIO, [**25**, **50**, **75**])) **20000.0** |

La salida correspondiente resulta

|  |
| --- |
| El salario mínimo es: **0.0**  El salario máximo es: **9765833.0**  Los cuartiles de salario son: [**18483.35** **32830.67** **52659.67**] |

**e. Presentar boxplot que muestren salario diferenciando por género y región**

El script de cálculo para este apartado es

|  |
| --- |
| # A2.e) Presentar boxplot de salario por género y región  filtro1 = datos\_csv[SEXO == **1**]  filtro2 = datos\_csv[SEXO == **2**]  plt.figure(figsize=(**8**, **6**))  plt.boxplot([filtro1['Salario'], filtro2['Salario']], labels=['Varones', 'Mujeres'])  plt.title("Boxplot de Salarios por Género")  plt.xlabel("Género")  plt.ylabel("Salario")  plt.show()  filtro1 = datos\_csv[REGION == **1**]  filtro2 = datos\_csv[REGION != **1**]  plt.figure(figsize=(**8**, **6**))  plt.boxplot([filtro1['Salario'], filtro2['Salario']], labels=['Montevideo', 'Interior'])  plt.title("Boxplot de Salarios por Región")  plt.xlabel("Región")  plt.ylabel("Salario")  plt.show() |

La salida correspondiente resulta

A picture containing text, screenshot, diagram, line

Description automatically generated

A picture containing text, screenshot, diagram, line

Description automatically generated

# ESTIMACIÓN DE PARÁMETROS

## DESEMPLEO

Como la variable desempleo extraída del archivo .CSV es una de tipo binario, entonces es pertinente tratar la tasa de desempleo anual como una variable de proporción

Donde Xi es una variable binaria (0, 1) extraída de la población de Bernoulli B(1, p) cuyo parámetro p es el porcentaje de éxitos en la población, luego por propiedad de media y varianza

Que la variable desempleo tenga distribución normal, implica que

Donde n es el tamaño de PEA y como n ≥ 30, el ET se estima según (Cordova, 2012)

Siendo p0 la proporción muestral de desempleados respecto a la PEA. Ahora a una confianza del 95 % (α = 0.05), el intervalo de confianza asociado a “p” será

Y el intervalo de confianza IC asociado a la variable desempleo se obtiene de multiplicar ICp por la población de la PEA en Uruguay que por dato es 1’757,161

**a. Estimar el desempleo del total de la población**

El script de cálculo para este apartado es

|  |
| --- |
| # B1) Estimar el desempleo del total de la población  **print**("**\n**El desempleo estimado es: ", int(td/**100**\***1757161**)) |

La salida correspondiente resulta

|  |
| --- |
| El desempleo estimado es: **132564** |

**a. Elabora intervalo de confianza con 95% de certeza para la variable desempleo.**

El script de cálculo para este apartado es

|  |
| --- |
| # B2) Elaborar IC para la variable desempleo al 95%  zo = stats.norm.ppf(**1** - **0.05**/**2**)  ET = (td/**100**\*(**1**-td/**100**)/PEA.sum())\*\***0.5**  LCi = td/**100**-ET\*zo  LCs = td/**100**+ET\*zo  **print**("**\n**El IC para el desempleo: ", [int(LCi\***1757161**),int(LCs\***1757161**)]) |

La salida correspondiente resulta

|  |
| --- |
| El IC para el desempleo: [**127046**, **138082**] |

# PRUEBA DE HIPÓTESIS

## DESEMPLEO

**a. Dada una tasa de desempleo en el 2021 de 7,0% (3). Con una certeza del 95%, ¿es correcto decir que la tasa de desempleo aumentó respecto del 2021?**

El estadístico de prueba en este caso corresponde a

Donde:

= Proporción muestral de desempleados

n = Tamaño muestral de la PEA

p = Proporción poblacional de desempleados

Luego se probará al 95%:

Ho: p ≥ 0.07, contra

H1: p < 0.07

Resultando la siguiente región de aceptación RA

El script de cálculo para este apartado es

|  |
| --- |
| # C1) Prueba de Hipótesis - Desempleo  zo = stats.norm.ppf(**0.05**)  ET = (**7**/**100**\*(**1**-**7**/**100**)/PEA.sum())\*\***0.5**  RAi = **7**/**100**+ET\*zo  **if** td >= **7**:  **print**("**\n**La tasa de desempleo aumentó respecto del 2021")  **else**:  **print**("**\n**La tasa de desempleo disminuyó respecto del 2021") |

La salida correspondiente resulta

|  |
| --- |
| La tasa de desempleo aumentó respecto **del** **2021** |

## SALARIO

El estadístico de prueba en este caso corresponde a

Donde:

= Media muestral para la variable salario en varones

= Media muestral para la variable salario en mujeres

n1 = Tamaño muestral para la variable salario en varones

n2 = Tamaño muestral para la variable salario en mujeres

μ1 = Media poblacional para la variable salario en varones

μ2 = Media poblacional para la variable salario en mujeres

S1 = estimador puntual de la desviación estándar para la variable salario en varones

S2 = estimador puntual de la desviación estándar para la variable salario en mujeres

Luego se probará al 99%:

Ho: μ1 – μ2 = 0, contra

H1: μ1 – μ2 ≠ 0

Resultando la siguiente región de aceptación RA

**a. A partir de los datos de la muestra, con una certeza del 99%, ¿hay diferencias en el salario promedio si distinguimos por género?**

El script de cálculo para este apartado es

|  |
| --- |
| # C2) Prueba de Hipótesis - Salario  filtro1 = datos\_csv[(PEA == **1**) & (DESEMPLEO == **0**) & (SEXO == **1**)]  filtro2 = datos\_csv[(PEA == **1**) & (DESEMPLEO == **0**) & (SEXO == **2**)]  zo = stats.norm.ppf(**1** - **0.01**/**2**)  n1 = len(filtro1)  n2 = len(filtro2)  S1 = np.std(filtro1['Salario'])  S2 = np.std(filtro2['Salario'])  ET = (S1\*\***2**/n1+S2\*\***2**/n2)\*\***0.5**  RAi = **0**-ET\*zo  RAs = **0**+ET\*zo  m1 = np.mean(filtro1['Salario'])  m2 = np.mean(filtro2['Salario'])  **if** ((m1-m2 >= RAi) & (m1-m2 <= RAs)):  **print**("**\n**Los salario de varones y mujeres son iguales")  **else**:  **print**("**\n**Los salario de varones y mujeres son distintos") |

La salida correspondiente resulta

|  |
| --- |
| Los salario de varones y mujeres son distintos |

## FUENTES

Banco Mundial. (2022). *Banco Mundial*. Obtenido de Banco Mundial: https://datos.bancomundial.org/indicator/SL.TLF.TOTL.IN?locations=UY

Cordova, M. (2012). *Estadística Descriptiva e Inferencial.* Lima: PUCP.

INE. (2023). *Instituto Nacional de Estadística*. Obtenido de https://www.ine.gub.uy/Anda5/index.php/catalog/730/get-microdata