

Practical1_GestionDatos_EstUnivariante

February 8, 2023

0.1 PRÁCTICA 1 - Gestión de Datos. Estadística Descriptiva Univariante.

Objetivos:

Manejo básico de estadística en Python:

- Gestión de datos
- Manejo de ficheros
- Análisis descriptivo de una y dos variables
- Búsqueda de relación entre variables
- Realización de análisis estadísticos y generación de informes

0.1.1 Instalamos las librerías necesarias

Primero, es necesario instalar las librerías `numpy`, `pandas`, `scipy`, `statsmodels`, `matplotlib` y `seaborn`

Si tenéis instalado `conda`:

```
conda install -c conda-forge numpy
conda install -c conda-forge pandas
conda install -c conda-forge scipy
conda install -c conda-forge statsmodels
conda install -c conda-forge matplotlib
conda install seaborn
```

Si vuestra instalación de Python es básica:

```
pip3 install numpy
pip3 install pandas
pip3 install scipy
pip3 install statsmodels
pip3 install matplotlib
pip3 install seaborn
```

Todos estos comandos los tenéis que ejecutar en vuestra Terminal, para que la instalación sea en vuestro ordenador y no en un ambiente creado.

```
[ ]: #Importamos las librerías que vamos a utilizar

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

0.1.2 Introducción a la estadística en Python: Conceptos Básicos

1. Creación manual de un fichero de datos (pd.DataFrame): Introducción de las variables Estatura (numérica) y Sexo (entero, 1=Hombre, 2=Mujer).

Estatura de 60 personas expresada en metros: 1.62 1.71 1.57 1.61 1.80 1.91 1.58 1.63 1.62 1.70 1.75 1.68 1.54 1.79 1.72 1.68 1.90 1.69 1.73 1.85 1.60 1.60 1.62 1.77 1.71 1.89 1.92 1.65 1.99 2.05 1.41 1.67 1.93 1.55 2.04 1.73 1.80 1.83 1.75 1.66 1.93 1.85 1.84 1.68 1.63 1.75 1.77 1.84 1.85 1.90 2.00 1.83 2.01 1.82 1.65 1.72 1.68 1.73 1.54 1.65

Sexos correspondientes a los datos anteriores: 2 1 2 1 1 2 2 2 2 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 1 1 1 2 1 2 2 2 1 2 2 2 2 1 2 2 1 1 2 2 2

Con los datos de Sexos, reemplazar el valor numérico por texto (1=Hombre; 2=Mujer) y crear un pd.DataFrame

```
[ ]: #1a. Creamos una lista de datos para la Estatura y el Sexo correspondiente
```

```
[ ]: #1b. Reemplazamos el valor numérico por texto
```

```
[ ]: #1c. Generamos un DataFrame con pd.DataFrame()
```

2. Análisis descriptivo / variables cuantitativas: Análisis descriptivo numérico y gráfico.

- Resumen estadístico (media, mediana, desviación estandar, quantiles...) y tabla de frecuencias.
- Hacer un gráfico de dispersión, un gráfico de caja y bigotes, y un histograma

```
[ ]: #2a. Realizamos el resumen estadístico. PISTA: hay una función que genera las
#     variables básicas que necesitamos.
```

```
[ ]: #2b. Generamos una tabla de frecuencias.
```

```
[ ]: #2c. Generamos un gráfico de dispersión.
```

```
[ ]: #2d. Generamos un gráfico de cajas y bigotes.
```

```
# BONUS: hay un tipo de gráfico que representa también la función de  
↪distribución. Intenta encontrarlo.
```

```
[ ]: #2e. Generamos un histograma.
```

3. Análisis descriptivo / variables cualitativa: Análisis descriptivo numérico y gráfico.

- Resumen estadístico (frecuencia) y tabla de frecuencias.
- Hacer un diagrama de barras y un diagrama de sectores

```
[ ]: #3a. Realizamos el resumen estadístico. PISTA: hay una función que genera las  
# variables básicas que necesitamos.
```

```
[1]: #3b. Generamos una tabla de frecuencias.
```

```
[ ]: #3c. Generamos un diagrama de barras.
```

```
[ ]: #3d. Generamos un diagrama de sectores.
```

4. Hallar la estatura media por sexo y generar los histogramas de alturas por sexo para compararlos.

- Previamente, manipula las opciones gráficas de la forma que se desee.

```
[ ]: #4a. Halla la estatura media por sexo
```

```
[ ]: #4b. Genera los histogramas de alturas por sexo y poder compararlos.
```

```
[ ]: #4c. Manipula el gráfico anterior en la forma que desees para poder ser  
↪presentado.
```

5. El dataset mpg contiene información sobre características de un conjunto de vehículos.

1. Para cargar el dataset, usad la siguiente función `sns.load_dataset('mpg')`
2. Haz una tabla de frecuencias, un diagrama de barras y un diagrama de sectores de la variable `model_year` (año de fabricación).
3. Haz una tabla de frecuencias para la variable `mpg`. ¿Qué porcentaje de coches consumen menos de 22 mpg (usa intervalos de 22)?
4. Haz el box-and-whisker plot de alguna de las variables cuantitativas y comenta sus características más importantes (concentraciones, asimetrías, atípicos...).

5. Haz un polígono de frecuencias RELATIVAS acumuladas de la variable **horsepower** (potencia - caballos).
6. Utilizando el polígono o diagrama de frecuencias relativas acumuladas, di aproximadamente la proporción de vehículos que tienen una potencia menor o igual a 100 CV.

```
[ ]: #5x. Carga el dataset mpg.
```

```
mpg = sns.load_dataset('mpg')
```

```
[ ]: #5a. Genera una tabla de frecuencias, un diagrama de barras y un diagrama de  
      ↳ sectores de la variable  
      # 'model_year' (año de fabricación)
```

```
[ ]: #5b. Genera una tabla de frecuencias de la variable 'mpg'.  
      # ¿Qué porcentaje de coches consumen menos de 22 mpg (usa intervalos de  
      ↳ 22)?
```

```
[ ]: #5c. Genera un box-and-whisker plot de alguna de las variables cuantitativas y  
      # comenta sus características más importantes (concentraciones, asimetrías,  
      ↳ atípicos...)
```

```
# BONUS: hay un tipo de gráfico que representa también la función de  
↳ distribución. Intenta encontrarlo.
```

```
[ ]: #5d. Genera un polígono de frecuencias RELATIVAS acumuladas de la variable  
      ↳ 'horsepower' (potencia - caballos).
```

```
[ ]: #5e. Utilizando la figura previa, di aproximadamente la proporción de vehículos  
      ↳ que tienen una  
      # potencia menor o igual a 100 CV.
```

Realizar un informe con los análisis realizados de cada uno de los apartados anteriores y exportarlo en un fichero .PDF El informe se realizará en un grupos de dos personas analizando los resultados obtenidos en cada uno de los apartados anteriores. Todas las gráficas deberán estar bien maquetadas: título, título en los ejes, ejes con un intervalo lógico y leyenda (si es el caso). El código empleado también se deberá incluir.

En el Moodle se encuentra un trabajo tipo para que tengáis de referencia. El archivo incluirá los nombres de todos autores en el siguiente formato:

Nombre del archivo: Practica1_ApellidosNombre1_ApellidosNombre2.pdf (extensión obligatoria en .pdf)

Se avisará en Moodle la fecha límite de entrega de este informe.