



CAMPUS PÚBLICO  
MARÍA ZAMBRANO  
SEGOVIA

---

**Universidad de Valladolid**

# Proyecto de la asignatura de Estadística

PROFESOR: CARLOS CALVO  
RAUL DESIDERIO MOTO BUAMBO  
RENZO SANDOVAL VILLANUEVA

## ÍNDICE

Resumen .....	2
1.Introducción .....	2
2. Metodología .....	2
2.1 Creación del conjunto de datos .....	2
2.2 Preparación y limpieza de datos.....	2
2.3 Análisis Descriptivo .....	3
2.4 Visualizacion de datos .....	3
2.5 análisis de una base de datos externa (mpg).....	3
3. RESULTADOS.....	3
4. CONCLUSIONES .....	5

## RESUMEN

Pandas y Seaborn facilitan el análisis, manipulación y visualización de datos.

## 1.INTRODUCCIÓN

En el ámbito del análisis de datos y la visualización en Python, dos bibliotecas fundamentales son Pandas y Seaborn. Ambas herramientas facilitan el procesamiento, análisis y representación gráfica de datos de manera eficiente y profesional.

**Pandas** es una biblioteca diseñada para trabajar con estructuras de datos en Python, permitiendo la manipulación y análisis de grandes volúmenes de información. Su principal estructura es el **DataFrame**, una tabla similar a una hoja de cálculo de Excel, donde los datos se organizan en filas y columnas.

**Seaborn** es una biblioteca basada en **Matplotlib** que facilita la creación de gráficos estadísticos con un diseño atractivo y optimizado. Permite generar visualizaciones complejas con pocas líneas de código, ideal para análisis exploratorio de datos.

**NumPy** se especializa en cálculos numéricos eficientes, mientras que **Matplotlib** facilita la creación de gráficos para visualizar datos.

## 2. METODOLOGÍA

### 2.1 CREACIÓN DEL CONJUNTO DE DATOS

- Se almacenan los datos en arrays (listas en Python) para facilitar su manipulación.
- Se incluyen dos listas: Estatura (valores numéricos en metros) y Sexo (valores numéricos 1 y 2).

### 2.2 PREPARACIÓN Y LIMPIEZA DE DATOS

- Se reemplazan los valores numéricos del sexo por etiquetas textuales (“H” y “M”).

- Se organizó la información en un DataFrame de Pandas

## 2.3 ANÁLISIS DESCRIPTIVO

Se calcularon medidas estadísticas (media, desviación, cuartiles, etc) con la función **describe()**.

Se generaron tablas de frecuencias para analizar la distribución de los datos.

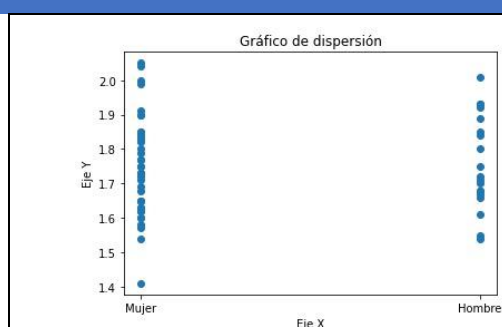
## 2.4 VISUALIZACION DE DATOS

- Gráficos de dispersión para observar la relación entre sexo y estatura.
- Diagramas de cajas para analizar la variabilidad de la estatura por sexo.
- Histogramas para observar la distribución de las alturas.
- Diagrama de barras y pastel para visualizar frecuencias de diferentes categorías.

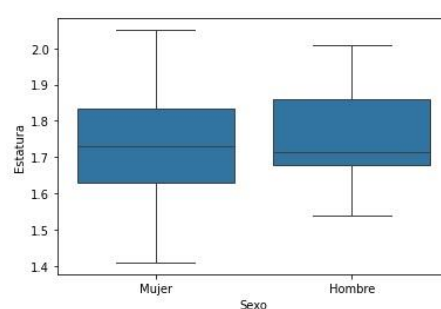
## 2.5 ANÁLISIS DE UNA BASE DE DATOS EXTERNA (MPG)

- Se cargó el dataset mpg de Seaborn.
- Se analizaron las distribuciones de **model\_year**, **mpg** y **horsepower** con gráficos y tablas de frecuencia.
- Se calculó la proporción de autos con consumo menor a 22 mpg.

## 3. RESULTADOS



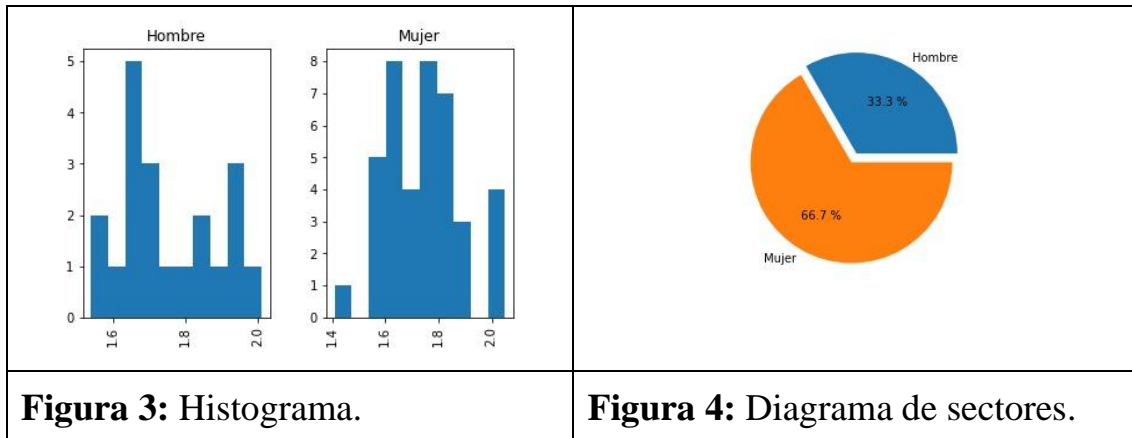
**Figura 1:** Gráfico de dispersión.



**Figura 2:** Gráfico de cajas y bigotes.

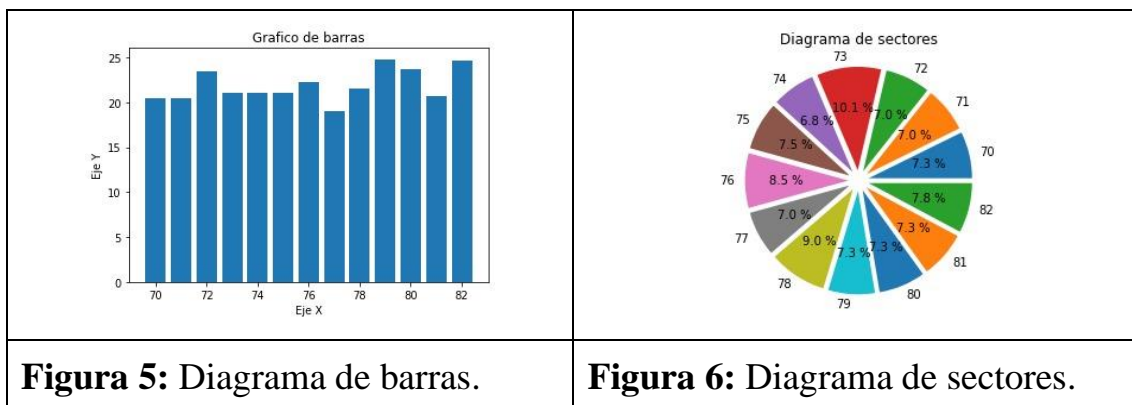
En la **figura 1** podemos observar los datos de las variables estatura, separados por el sexo, siendo hombre y mujer.

En la **figura 2** se puede apreciar tanto mediana como cuartiles de ambas variables, encontrándose en el eje **X** el sexo y en el eje **Y** la estatura.



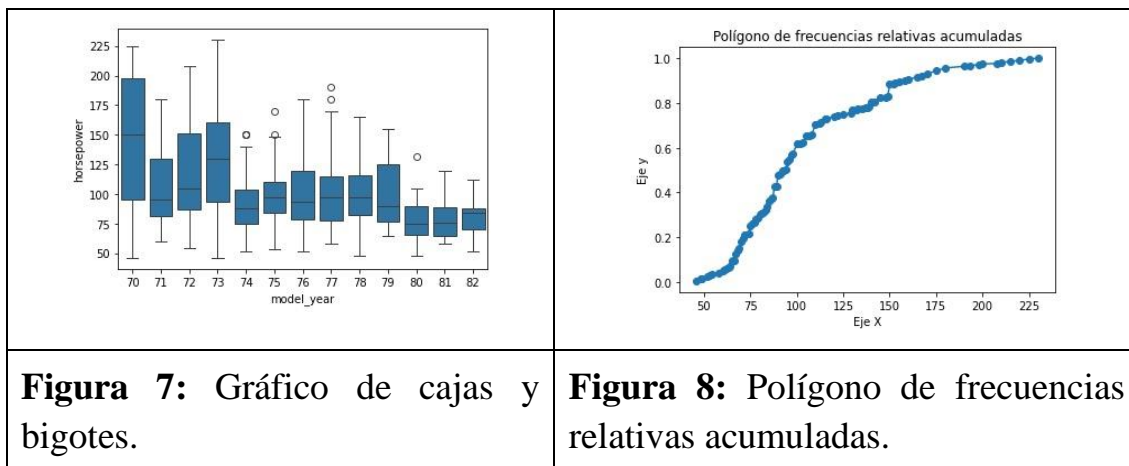
La **figura 3** muestra la información agrupada por intervalos sobre la estatura y separados por el sexo, en el eje X se encuentran los intervalos solapados de clase (estatura) y en el eje Y la frecuencia del intervalo.

En la **figura 4** tenemos un diagrama que muestra por porcentajes la cantidad tanto de hombres como de mujeres.



En la **figura 5**, se encuentra un diagrama que muestra los resultados por barras de la variable “model\_year” en el eje X, para poder distinguir mejor las barras, al eje Y le hemos asignado la variable “aceleración”.

La **figura 6** muestra por sectores y porcentajes la cantidad de datos de la variable “model\_year”.



En la **figura 7** muestra datos de las variables “model\_year” (eje X) y “horsepower” (eje Y), y encontramos tanto cuartiles como mediana.

La **figura 8** proporciona datos sobre la variable “horsepower”, separados por intervalos en el eje X y su frecuencia mostrada en el eje Y.

#### 4. CONCLUSIONES

En cuanto a los datos del primer ejercicio que trata sobre la estatura y sexo de un grupo de personas, al estudiar las gráficas del **boxplot** y los diagramas de barras, reflejando más variabilidad en cuanto a las estaturas entre hombre que entre mujeres, también en el diagrama de sectores, se puede observar que hay más mujeres que hombres, y en general que las mujeres tienen una altura superior a la de los hombres.

Para el análisis del segundo ejercicio que trata sobre un listado de coches, podemos ver en el diagrama de sectores la cantidad de coches (por porcentajes) que se han construido a lo largo de diferentes años, o por ejemplo una distribución de polígono de frecuencias de los mismos coches, pero distinguiéndose por sus caballos.