

Practical 3

AIM:

Provide summary statistics for a dataset with numeric variables grouped by one of the qualitative variables.

write a python program to display some basic statistical details like percentile, mean, standard deviation , etc.

REQUIREMENT:

Anaconda Installer

Windows10OS

JupyterNotebook

THEORY:

Step 1: Summary Statistics

Provide summary statistics such as **mean, median, mode, and standard deviation**.

Mean:

- Mean is the **average value** of the dataset
 - Formula:
Mean = (Sum of all values) / (Total number of values)
-

Median:

- Median is the **middle value** of a **sorted dataset**
 - Arrange data in **ascending order** before finding median
-

Mode:

- Mode is the **most frequently occurring value** in the dataset
-

Example:

Data: 36, 40, 32, 42, 30

- **Mean = $(36 + 40 + 32 + 42 + 30) / 5 = 36$ kg**
 - Arrange data: 30, 32, 36, 40, 42
 - **Median = 36 kg**
 - **Mode = No mode (no value repeats)**
-

Standard Deviation:

- Measures how much data values **deviate from the mean**
- It is the **square root of variance**
- Helps to understand **data spread or variation**

Standard Deviation Formula:

$$\sigma = \sqrt{\frac{(x_i - \mu)^2}{N}}$$

Where:

- σ = Standard Deviation
- x_i = Each data value
- μ = Mean of the dataset
- N = Total number of values

LibrariesUsed:

Pandas

Numpy

Seaborn

CONCLUSION:

From this experiment, we learned how to calculate **mean, median, and mode** to understand the central tendency of data. These measures help in summarizing data and analyzing its distribution effectively..