**Name : Adhik Baba Sarak**

**PRN : 123B2F152**

**Assignment No. 2**

**Problem Statement:**

**Write a python script to find basic descriptive statistics using summary, quartile function, etc on iris datasets.**

**Objective:** The objective of this assignment is to develop a Python script to compute and display basic descriptive statistics for the Iris dataset. This script will utilize functions such as describe() to summarize key statistics (mean, median, standard deviation, etc.) and calculate quartiles for each feature (sepal length, sepal width, petal length, petal width). The script will help in understanding the central tendency, spread, and distribution of the data.

**Prerequisite :**

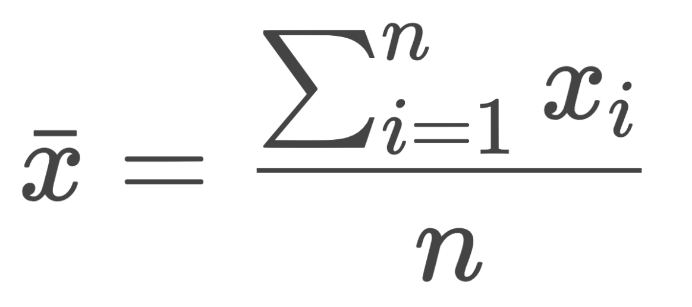
1. A Python environment set up with libraries like pandas,Numpy,Matplotlib.
2. Internet connection (for downloading datasets from the web).
3. **Descriptive statistics concepts** such as mean, median, mode, variance, standard deviation, quartiles, etc.
4. A working environment such as Jupyter Notebook to run Python scripts.

**Theory :**

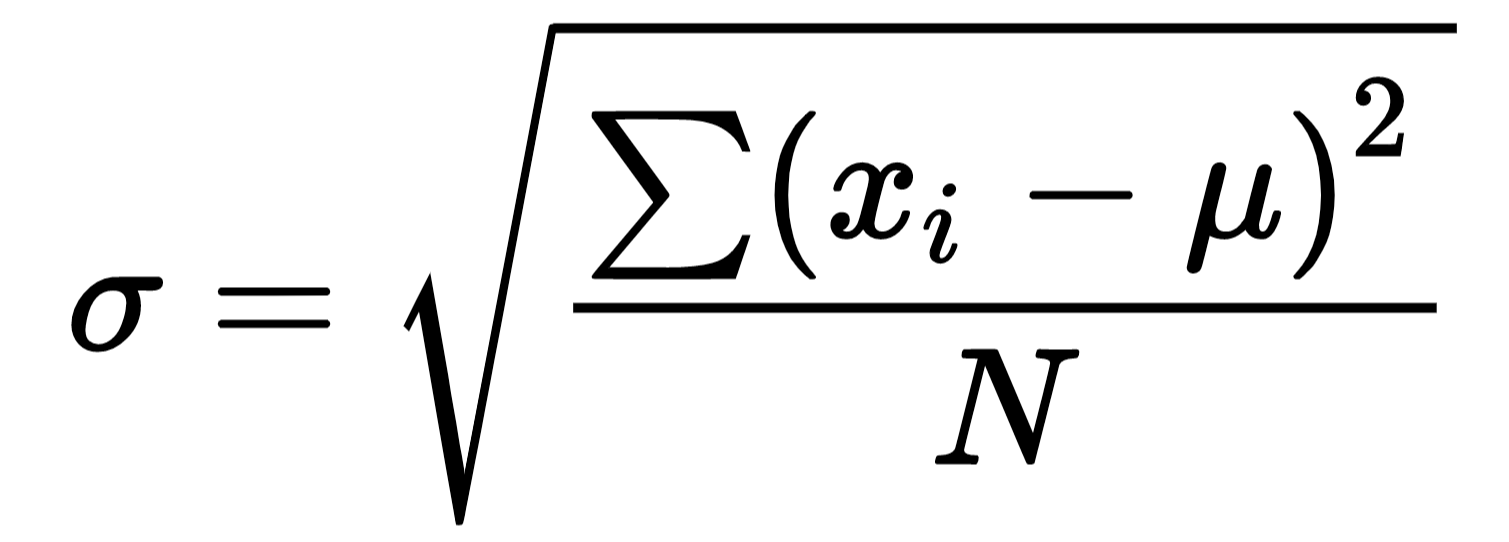
Descriptive statistics provide a summary of the essential aspects of a dataset, helping us understand the data's central tendency, dispersion, and overall distribution.

**Key Concepts in Descriptive Statistics:**

1. Mean (Average): The mean is the sum of all values in a dataset divided by the number of values. It gives an indication of the central value of the data.

****

1. Median: The median is the middle value of a dataset when the data is ordered from smallest to largest. It provides a measure of central tendency that is less sensitive to outliers than the mean.
2. Mode: The mode is the value(s) that appear most frequently in a dataset. A dataset can have no mode, one mode, or multiple modes.
3. Standard Deviation (SD): Standard deviation measures the spread or dispersion of the data around the mean. A lower SD indicates that the data points are closer to the mean, while a higher SD indicates more variability.

****

1. Variance: Variance is the average of the squared differences from the mean and provides a measure of data dispersion. It is the square of the standard deviation.
2. Minimum and Maximum: These values represent the smallest and largest data points in a dataset, respectively, giving the range of the data.
3. Quartiles: Quartiles divide the data into four equal parts, helping us understand the distribution of the data:

**Algorithm**

Step 1: Load the dataset:

* Load the Iris dataset using the sklearn.datasets.load\_iris() function.
* Create a DataFrame using pandas to store the Iris features (sepal length, sepal width, petal length, petal width).

Step 2: Calculate the mean:

* For each feature (sepal length, sepal width, petal length, petal width), compute the mean using the mean() function from pandas.

Step 3: Calculate the median:

* For each feature, compute the median using the median() function from pandas.

Step 4: Calculate the mode:

* For each feature, find the mode using the mode() function from pandas.

Step 5: Generate summary statistics:

* Use describe() to get summary statistics such as count, mean, std (standard deviation), min (minimum), max (maximum), and quartiles (Q1, Q3).

Step 6: Calculate quartiles:

* Use quantile(0.25) to get Q1 (the 25th percentile).
* Use quantile(0.75) to get Q3 (the 75th percentile).

Step 7: Print results:

Print the calculated mean, median, mode, and other summary statistics including the quartiles.

**References :**

* Descriptive Statistics-https://rfaqs.com/descriptive-statistics-in-python/
* Iris dataset-https://www.kaggle.com/datasets/vikrishnan/iris-dataset
* Pandas Documentatio**n**

**Conclusion**

In this task we conducted the computation of some basic descriptive statistics using the Iris dataset, a renowned dataset in machine learning and statistics. With the aid of python scripts using libraries like Pandas and Scikit-learn we were able to quickly calculate some important statistic measures such as resolving mean, median, mode, quartiles and total statistics summary of the four features of the Iris dataset, which include sepalLength, sepalWidth, petalLength and petalWidth.