Department of Engineering Sciences and Technology,

Second Year Btech in Computer Science

Project Based Learning-Python

<u>Assignment - 12</u>

Name - Paritosh kolwadkar

SRN – 31231313

Roll no – 39

Batch – D2

Problem statement : **Write a program to create a NumPy array and calculate its mean, median, standard deviation, and variance. Use advanced statistical methods to compute percentiles and correlation coefficients.**

Pre-requisites: Familiarity with NumPy for array creation and statistical operations.

Understanding of statistical terms such as mean, median, standard deviation, variance, percentiles, and correlation coefficients.

Code:

```python
# Import NumPy
import numpy as np

# Create a 1D NumPy array
array = np.array([10, 20, 30, 40, 50, 60, 70, 80, 90, 100])

print("Array:")
print(array)
```

```python
# Statistical computations
mean_value = np.mean(array)
median_value = np.median(array)
std_deviation = np.std(array)
variance_value = np.var(array)

print("\nStatistical Computations:")
print(f"Mean: {mean_value}")
print(f"Median: {median_value}")
print(f"Standard Deviation: {std_deviation}")
print(f"Variance: {variance_value}")

# Advanced statistical methods
percentile_25 = np.percentile(array, 25)  # 25th percentile
percentile_50 = np.percentile(array, 50)  # Median
percentile_75 = np.percentile(array, 75)  # 75th percentile

print("\nPercentiles:")
print(f"25th Percentile: {percentile_25}")
print(f"50th Percentile (Median): {percentile_50}")
print(f"75th Percentile: {percentile_75}")

# Create another array for correlation computation
array2 = np.array([15, 25, 35, 45, 55, 65, 75, 85, 95, 105])

# Compute correlation coefficient
correlation_matrix = np.corrcoef(array, array2)

print("\nCorrelation Coefficient Matrix:")
```

```
print(correlation_matrix)
```

Explanation :

Array Creation:

- **A 1D array `array` is created using `np.array`.**

Basic Statistics:

- `np.mean(array)`: **Computes the mean (average).**
- `np.median(array)`: **Finds the middle value when the data is sorted.**
- `np.std(array)`: **Calculates the standard deviation (spread of the data).**
- `np.var(array)`: **Computes the variance (measure of data variability).**

Percentiles:

- `np.percentile(array, p)`: **Computes the ppp-th percentile. For example, the 25th, 50th, and 75th percentiles divide the data into quarters.**

Correlation Coefficients:

- `np.corrcoef(array1, array2)`: **Calculates the correlation matrix between two arrays, measuring how strongly the variables are related.**

Output:

Array:

[10 20 30 40 50 60 70 80 90 100]

Statistical Computations:

Mean: 55.0

Median: 55.0

Standard Deviation: 28.722813232690143

Variance: 825.0

Percentiles:

25th Percentile: 32.5

50th Percentile (Median): 55.0

75th Percentile: 77.5

Correlation Coefficient Matrix:

[[1. 1.]

 [1. 1.]]

Output Explained:

Basic Statistics:

- **Mean: $(10+20+\ldots+100)/10=55$ (10 + 20 + \ldots + 100) / 10 = 55$(10+20+\ldots+100)/10=55$**
- **Median: The middle value of sorted data: 55.**
- **Standard Deviation: Measures the spread of data points.**
- **Variance: Square of the standard deviation.**

Percentiles:

- **25th Percentile: Value below which 25% of data lies: 32.5.**
- **50th Percentile: The median: 55.**
- **75th Percentile: Value below which 75% of data lies: 77.5.**

Correlation Coefficient Matrix:

- **Diagonal values (1.0) indicate perfect correlation between each array and itself.**
- **Off-diagonal values (1.0) indicate perfect correlation between `array` and `array2`.**