Experiment 20

Aim: - Write a Python program to calculate the mean, median, standard deviation, variance, and correlation coefficients of a given array.

Theory

Statistical Operations on an Array

Statistical operations help summarize data and understand its distribution. NumPy provides functions to compute these metrics efficiently.

1. Mean (Average)

The mean is the sum of all elements divided by the total number of elements:

$$\text{Mean} = \frac{\sum X_i}{N}$$

mean_value = np.mean(data)

2. Median

The median is the middle value of a sorted dataset. If there are an even number of elements, it is the average of the two middle values.

median_value = np.median(data)

3. Standard Deviation

Standard deviation measures the spread of data from the mean.

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

std_dev = np.std(data)

4. Variance

Variance represents the average squared deviation from the mean.

$$\text{Variance} = \sigma^2$$

variance = np.var(data)

5. Correlation Coefficient

The correlation coefficient measures the relationship between two datasets. It ranges from -1 to 1.

```
correlation_matrix = np.corrcoef(x, y)
```

Algorithm: Statistical Operations

Step 1: Import NumPy

- Import the NumPy library to perform calculations.

Step 2: Create a Data Array

- Define a dataset using np.array().

Step 3: Compute Statistics

- Calculate the mean using np.mean().
- Calculate the median using np.median().
- Calculate the standard deviation using np.std().
- Calculate the variance using np.var().

Step 4: Compute Correlation Coefficients

- Define two related datasets.
- Use np.corrcoef() to compute the correlation matrix.

Step 5: Display Results

- Print all calculated values.

Code

```
import numpy as np


# Creating an array
data = np.array([10, 20, 30, 40, 50, 60, 70, 80, 90, 100])


# Calculating statistical measures
mean_value = np.mean(data)  # Mean
median_value = np.median(data)  # Median
std_dev = np.std(data)  # Standard Deviation
variance = np.var(data)  # Variance
```

```
# Creating two related datasets for correlation
x = np.array([1, 2, 3, 4, 5])
y = np.array([2, 4, 6, 8, 10])
correlation_matrix = np.corrcoef(x, y)  # Correlation Coefficient


# Printing results
print("Mean:", mean_value)
print("Median:", median_value)
print("Standard Deviation:", std_dev)
print("Variance:", variance)
print("Correlation Coefficients:")
print(correlation_matrix)
```

Ouput