**Student ID:** 21127690

**Name:** Ngô Nguyễn Thanh Thanh

**Report: Analysis of Basic Statistical Functions Using NumPy**

**I. Evaluation summary:**

|  |  |  |
| --- | --- | --- |
| **Function** | **Requirement Met(%)** | **Notes** |
| Mean | 100% |  |
| Median | 100% |  |
| Variance | 100% |  |
| Standard Deviation | 100% |  |
| Order Statistics(min, max,range) | 100% |  |
| Correlation | 100% |  |

**II. List of funtion:**

*Including image proof below*

**1. Mean:**

**A computer screen shot of text

Description automatically generated**

**A black screen with white text

Description automatically generated**

**2. Median**

**A screenshot of a computer program

Description automatically generated**

**A black screen with white text

Description automatically generated**

**3. Variance**

**A computer screen shot of text

Description automatically generated**

**A black screen with white text

Description automatically generated**

**4. Standard Deviation**

**A computer screen shot of text

Description automatically generated**

**A black screen with white text

Description automatically generated**

**5. Min**

**A screenshot of a computer program

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**6.Max**

**A screenshot of a computer program

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**7. Range**

**A computer screen with colorful text

Description automatically generated**

**A black screen with white numbers and symbols

Description automatically generated**

**8. Correlation**

**A computer screen shot of text

Description automatically generated**

**A number on a black background

Description automatically generated**

**Notes:**

* All of the functions mentioned above have two versions: one for handling regular numbers and another for handling non-numeric values.
* In summary, NumPy's functions automatically handle non-numeric values by excluding them from calculations where appropriate. This behavior ensures that the statistical analysis is performed accurately without being affected by NaN values in the input data.

**III. Function Summaries and Implementation**

***Desciption methods of Numpy library and examples:***

*Because the task requires utilizing NumPy to compute basic statistical values, therefore the code only calls NumPy library functions such as np.median, np.mean, etc. The explanation below elucidates how the statistical values are computed behind the scenes by the functions in the NumPy library.*

* **Mean:** Calculates the arithmetic mean of a dataset. It is implemented using the formula sum of elements divided by the total number of elements.
* **Median:** Finds the middle value of a dataset. It is implemented by sorting the dataset and finding the middle value or the average of the two middle values.

**EX:**

A math problem with numbers and lines

Description automatically generated with medium confidence

* **Variance:** Measures the spread of data points around the mean. It is calculated by finding the average of the squared differences from the mean.
* **Standard Deviation**: Represents the average deviation of data points from the mean. It is the square root of the variance.

**EX:**

A math equation with numbers and symbols

Description automatically generated

* **Order Statistics**: Includes functions like min and max, which find the minimum and maximum values in a dataset, respectively.
* **Range:** Indicates the spread or difference between the maximum and minimum values in a dataset. It provides insight into the variability of the data.

**EX:**

A screen shot of a computer

Description automatically generated

Range of X = 99 – 14 = 85 (in the scope of full array)

Range X (axis = 0) = [85 29] (in the first colum the range is 99 -14= 85, in the second colum the range is 96 – 67 = 29)

Range X (axis = 1) = [82 36 13 14 12] (similar to above but with each row)

96 – 14 = 82, 82 – 46 = 36, 80 – 67 = 13, 91 – 77 = 14, 99 – 87 = 12)

* **Correlation:** Measures the strength and direction of the linear relationship between two variables. It is calculated using covariance and standard deviation.

**EX:**

A black background with white text

Description automatically generated

A black background with white text

Description automatically generated

After substituting the values above into the formula, we obtain the result: -0.99999

Top of Form

**V. The results:**

***Full image proof:***

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**