

Information Analysis and Visualisation

Laboratory Session 5

Statistical Analysis

Statistical Analysis is a key topic in Mathematics and Data Science. This laboratory session is focused on a range of Statistics-related problems. We will be using the Python libraries NumPy and pandas, which provide the data structures and the functionality needed for solving statistical problems. Additionally, we will be using the Matplotlib library to visualise the datasets.

Task 1

Initialise an array containing 20 random values from the interval [0 .. 100]. Visualise the values of the data points and plot the dataset by using a Line Chart.

Libraries:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

Task 2

By using functions from the NumPy library, find the following statistical measures:

- Mean value
- Median value
- Range
- Standard deviation

Task 3

Initialise a pandas Data Frame with the values from the previously initialised array. Print both data structures and observe any differences in the output.

Task 4

By using the 'describe' method from the pandas library provide the following information for the generated dataset:

- Mean
- Lower quartile
- Median
- Upper quartile

Task 5

Calculate the Interquartile range for the dataset.