Aim: To study the measures central tendency and measures of dispersion for univariate data.

Theory:

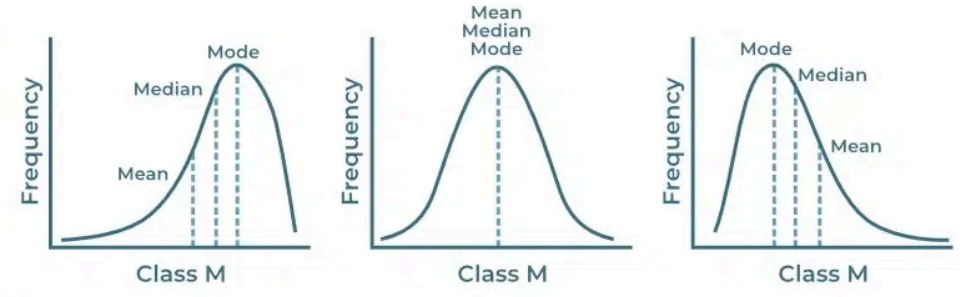
Univariate Data:

Univariate data refers to a type of data in which each observation or data point corresponds to a single variable. In other words, it involves the measurement or observation of a single characteristic or attribute for each individual or item in the dataset. Analyzing univariate data is the simplest form of analysis in statistics.

Heights (in cm)	164	167.3	170	174.2	178	180	186

Measures of Central Tendency:

Central Tendencies in Statistics are the numerical values that are used to represent mid-value or central value in a large collection of numerical data. These obtained numerical values are called central or average values in Statistics. A central or average value of any statistical data or series is the value of that variable that is representative of the entire data or its associated frequency distribution. Such a value is of great significance because it depicts the nature or characteristics of the entire data, which is otherwise very difficult to observe.



1. **Mean**

   A mean is a quantity representing the "center" of a collection of numbers and is intermediate to the extreme values of the set of numbers.

$$m = \frac{\text{sum of the terms}}{\text{number of terms}}$$

$$\bar{x} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

2. **Median**

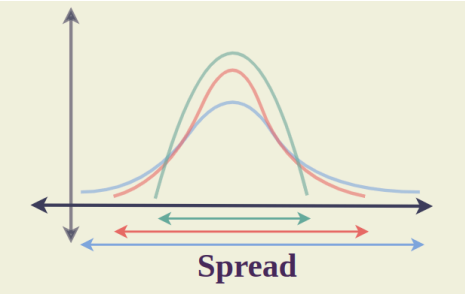
   The median of a set of numbers is the value separating the higher half from the lower half of a data sample, a population, or a probability distribution. For a data set, it may be thought of as the “middle" value.

$$\text{med}(x) = x_{(n+1)/2} \quad \text{(if n is odd)}$$

$$\text{med}(x) = \frac{x_{(n/2)} + x_{((n/2)+1)}}{2}$$ (if n is even)

3. **Mode**
   A mode is defined as the value that has a higher frequency in a given set of values. It is the value that appears the most number of times.

<u>Measures of Dispersion:</u>



Dispersion in statistics is a way to describe how spread out or scattered the data is around an average value. It helps to understand if the data points are close together or far apart. Dispersion shows the variability or consistency in a set of data. There are different measures of dispersion like range, variance, and standard deviation. Measures of Dispersion measure the scattering of the data. It tells us how the values are distributed in the data set. In statistics, we define the measure of dispersion as various parameters that are used to define the various attributes of the data.

1. **Range**
   It is defined as the difference between the largest and the smallest value in the distribution. A higher value of range implies higher variation in the data set. One drawback of this measure is that it only takes into account the maximum and the minimum value. They might not always be the proper indicator of how the values of the distribution are scattered.

   Range = **H**ighest value - **L**owest value

2. **Standard Deviation**
   It is the square root of the arithmetic average of the square of the deviations measured from the mean. In statistics, the standard deviation is a measure of the amount of variation of the values of a variable about its mean. A low standard deviation indicates that the values tend to be close to the mean of the set, while a high standard deviation indicates that the values are spread out over a wider range.

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

3. **Variance**
   In probability theory and statistics, variance is the expected value of the squared deviation from the mean of a random variable. The standard deviation is obtained as the square root of the variance.

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

4. **Quartiles**

In statistics, quartiles are a type of quantiles which divide the number of data points into four parts, or quarters, of more-or-less equal size. The data must be ordered from smallest to largest to compute quartiles; as such, quartiles are a form of order statistic.
Q2 = Median
Q1 = Q1 is the median of the lower half of the data (excluding the median if n is odd).
Q3 = Q3 is the median of the upper half of the data (excluding the median if n is odd).

Code:

```
import pandas as pd

file_path = 'Accumulative_distribution.csv'
# Load the data
data = pd.read_csv(file_path)

# Selecting relevant numerical columns for analysis
numerical_columns = ['Distance']

# Calculating the required statistical measures
results = {}
for column in numerical_columns:
    stats = {
        "Mean": data[column].mean(),
        "Median": data[column].median(),
        "Mode": data[column].mode().iloc[0] if not data[column].mode().empty else None,
        "Minimum": data[column].min(),
        "Maximum": data[column].max(),
        "Range": data[column].max() - data[column].min(),
        "Standard Deviation": data[column].std(),
        "Variance": data[column].var(),
        "Quartiles": {
            "Q1": data[column].quantile(0.25),
            "Q2 (Median)": data[column].quantile(0.5),
            "Q3": data[column].quantile(0.75)
        }
    }
    results[column] = stats

# Display the results
for column, stats in results.items():
    print(f"--- {column} ---")
    for measure, value in stats.items():
```

```
    if isinstance(value, dict):
        print(f"  {measure}:")
        for q, q_value in value.items():
            print(f"    {q}: {q_value}")
    else:
        print(f"  {measure}: {value}")
```

```
--- Distance ---
  Mean: 42.227294339303455
  Median: 39.10371649
  Mode: 0.993842499
  Minimum: 0.993842499
  Maximum: 114.943724
  Range: 113.94988150100001
  Standard Deviation: 32.33535637753354
  Variance: 1045.575272062099
  Quartiles:
    Q1: 8.79283179175
    Q2 (Median): 39.10371649
    Q3: 70.26854610000001
```

Dataset - Insects Flight Dynamics 

Conclusion:
Thus, the mean, median, mode, range, standard deviation, variance, quartiles were calculated for the given dataset.