

Q1.

Three measures of Central Tendency are :

- Mean , Median and Mode.

Q2.

Mean :

- is calculated by adding up all the values in a dataset and then dividing by the number of values.
- It is useful when the data is roughly symmetric and does not have extreme values that could skew the result.

Median :

- The median is the middle value of a dataset when it is ordered from least to greatest.
- The median is useful when the dataset has extreme values

Mode :

- The mode is the value that appears most frequently in a dataset.
- The mode is useful for categorical data or when identifying the most common value is important.

Q3.

```
In [1]: height = [178,177,176,177,178.2,178,175,179,180,175,178.9,176.2,177,172.5,178,17
```

```
In [2]: import numpy as np
np.mean(height)
```

```
Out[2]: 177.01875
```

```
In [3]: np.median(height)
```

```
Out[3]: 177.0
```

```
In [4]: import statistics
```

```
In [5]: statistics.mode(height)
```

```
Out[5]: 178
```

Q4.

```
In [6]: data = [178,177,176,177,178.2,178,175,179,180,175,178.9,176.2,177,172.5,178,176.  
np.std(data)
```

```
Out[6]: 1.7885814036548633
```

Q5.

Range :

- Range = Maximum Value - Minimum Value
- Consider the following dataset of exam scores - {60, 70, 75, 80, 90}. The range is $90 - 60 = 30$.

Variance :

- Variance measures the average squared deviation of each data point from the mean of the dataset.
- Using the exam scores dataset {60, 70, 75, 80, 90}, suppose the mean is calculated to be 75. The variance would be 130.

Standard Deviation :

- The standard deviation is the square root of the variance.
- Using the variance example, the standard deviation would be 11.40.

Q6.

A Venn diagram is a visual representation that illustrates the relationships between sets. It consists of overlapping circles, each representing a specific set, and the overlapping areas represent the elements that are common to those sets.

Q7.

$$i = \{2\}$$

$$ii = \{0, 2, 3, 4, 5, 6, 7, 8, 10\}$$

Q8.

Skewness is a statistical measure that describes the asymmetry or lack of symmetry in a distribution of data.

Q9.

Mean > Median.

Q10.

Covariance :

- Covariance is a measure of how much two variables change together. It indicates the direction of the linear relationship between variables, whether positive or negative.
- Covariance is used to understand the direction of the relationship between two variables.

Correlation :

- Correlation is a standardized measure that indicates the strength and direction of a linear relationship between two variables.
- Correlation is widely used because it provides a standardized measure that is independent of the scale of the variables.

Q11.

summation of $x(i)$ / n from $i = 1$ to n

Eg :

- exam scores: {80, 75, 90, 85, 88}.

Using the formula , we get sample mean as 83.6.

Q12.

Mean = Median = Mode. All are perfectly at the centre.

Q13.

They both differ in their calculation , interpretation and scale.

Q14.

Central tendency measures, such as the mean, are particularly sensitive to outliers, while dispersion measures, such as the range, variance, and standard deviation, can be influenced by the presence of extreme values.

Eg :

- exam scores: {75, 80, 85, 90, 95, 200}. The majority of scores fall in the range of 75 to 95, but there is one extreme outlier (200).
- By calculating mean , we get 104.17
- Median , which is 87.5
- Clearly says that median is less affected by the extreme value when compared to mean.

Dispersion :

Eg : dataset {10, 15, 20, 25, 30, 100}. The majority of values are between 10 and 30, but the outlier (100) significantly influences measures of dispersion.

- Range = $100 - 10 = 90$
- The range is larger due to the presence of the outlier.
- Using the formula for variance and standard deviation, you will find that these measures are also larger because the squared differences from the mean are larger with the outlier.