

Q1. The three measures of central tendency are:

1. Mean: It is the arithmetic average of a dataset and is calculated by summing all the values and dividing by the number of values. 2. Median: It is the middle value of a dataset when it is arranged in ascending or descending order. If there is an even number of data points, the median is the average of the two middle values. 3. Mode: It is the value that appears most frequently in a dataset. A dataset can have no mode, one mode (unimodal), or multiple modes (multimodal).

Q2. The key differences and uses of mean, median, and mode for measuring central tendency are as follows:

1. Mean: Provides the average value and is sensitive to extreme values (outliers). It is commonly used when data is symmetrically distributed. 2. Median: Represents the middle value and is less affected by extreme values. It is suitable for skewed data or data with outliers. 3. Mode: Represents the most frequent value and is useful for categorical or discrete data.

Q3. Measure the three measures of central tendency for the given height data:
[178,177,176,177,178.2,178,175,179,180,175,178.9,176.2,177,172.5,178,176.5]

Mean: Approximately 177.6 Median: 177.9 Mode: 178 (appears most frequently)

Q4. Find the standard deviation for the given data:
[178,177,176,177,178.2,178,175,179,180,175,178.9,176.2,177,172.5,178,176.5]

The standard deviation for the given data is approximately 1.75.

Q5. How are measures of dispersion such as range, variance, and standard deviation used to describe the spread of a dataset? Provide an example.

In []: 1. Range: It measures the difference between the maximum and minimum values in a dataset.
2. Variance: It measures the average of the squared differences from the mean.
3. Standard Deviation: It is the square root of the variance and provides a measure of the spread of the data.
Example: If you have test scores for two classes:
Class A: [90, 92, 88, 85, 94]
Class B: [78, 85, 82, 88, 90]
Class A has a higher mean, but Class B has a smaller standard deviation, indicating that Class B's scores are more consistent.

Q6. What is a Venn diagram?

A Venn diagram is a visual representation of the relationships between two or more sets. It consists of overlapping circles or ellipses, each representing a set, and the areas of overlap represent elements that belong to multiple sets. Venn diagrams are often used in set theory and statistics to illustrate the intersections and unions of sets.

Q7. For the two given sets A = (2,3,4,5,6,7) & B = (0,2,6,8,10). Find: (i) A ∩ B (ii) A ∪ B

```
In [ ]: (i) A ∩ B (intersection) = {2, 6}
(ii) A ∪ B (union) = {0, 2, 3, 4, 5, 6, 7, 8, 10}
```

Q8. What do you understand about skewness in data?

Skewness in data refers to the lack of symmetry in the data's distribution. If data is right-skewed, it means that the right tail (larger values) is longer or more spread out than the left tail (smaller values). Right-skewed data has a long tail on the right side.

Q9. If a data is right skewed then what will be the position of median with respect to mean?

In a right-skewed distribution, the median is typically less than the mean. This is because the long tail on the right side pulls the mean in that direction, making it greater than the median.

Q10. Explain the difference between covariance and correlation. How are these measures used in statistical analysis?

Covariance measures the degree to which two variables change together, indicating the direction of the linear relationship. Correlation, on the other hand, is a standardized measure of the strength and direction of the linear relationship between two variables. It ranges from -1 (perfect negative correlation) to 1 (perfect positive correlation).

Both covariance and correlation are used in statistical analysis to understand relationships between variables. Correlation is preferred as it is not affected by the scale of the variables, making it easier to interpret.

Q11. What is the formula for calculating the sample mean? Provide an example calculation for a dataset.

. The formula for calculating the sample mean (\bar{x}) is:

$$\bar{x} = (\sum x) / n$$

Example: For a dataset {12, 15, 18, 21, 24}, the mean is calculated as $(12+15+18+21+24)/5 = 18$.

Q12. For a normal distribution data what is the relationship between its measure of central tendency?

For a normal distribution:

The mean, median, and mode are all equal and located at the center of the distribution. The distribution is perfectly symmetric, with the mean, median, and mode all at the same point.

Q13. How is covariance different from correlation?

Covariance measures the degree and direction of the linear relationship between two variables, but it is not standardized and depends on the scales of the variables.

Correlation standardizes the measure, making it independent of scale and providing a value between -1 and 1, indicating the strength and direction of the linear relationship.

Q14. How do outliers affect measures of central tendency and dispersion? Provide an example.

Outliers can significantly affect measures of central tendency and dispersion:

Mean: Outliers can pull the mean in their direction, making it unrepresentative of the majority of the data. Median: The median is less affected by outliers since it only considers the middle value(s). Standard Deviation: Outliers can increase the standard deviation by increasing the dispersion. Example: In a dataset of incomes, a single extremely high income (outlier) can greatly increase the mean income while the median remains largely unaffected, giving a misleading representation of typical income.