

# Statistics

## Basic-2 Assignment

### Questions

### Assignment

Q1. What are the three measures of central tendency?

**Ans - Mean, median and mode.**

Q2. What is the difference between the mean, median, and mode? How are they used to measure the central tendency of a dataset?

**Ans - Mean: It is the average of all values in the dataset. It is sensitive to outliers and works best with symmetrically distributed data.**

**Median: It is the middle value when the dataset is arranged in ascending or descending order. It is less affected by outliers and suitable for skewed data.**

**Mode: It is the value that appears most frequently in the dataset. It can be used with any type of data, including categorical data.**

**Each measure helps describe the central value of a dataset, and the choice depends on the data's distribution and the presence of outliers.**

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]

**Ans - Mean: 177.77**

**Median: 177**

**Mode: 178**

**For the given height data, the mean height is approximately 177.77, the median height is 177, and the mode is 178. These measures help describe the central tendency of the dataset.**

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]

**Ans - Standard deviation is 1.77**

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

Q6. What is a Venn diagram?

**Ans - A Venn diagram is a graphical representation of sets using overlapping circles or ellipses. It is used to show the relationships and commonalities between different sets. Each circle in the diagram represents a set, and the overlapping regions represent the elements that belong to multiple sets. It helps visualize the intersections and differences between sets and is commonly used in mathematics, logic, statistics, and other fields to illustrate set theory concepts and logical relationships.**

Q7. For the two given sets  $A = \{2,3,4,5,6,7\}$  &  $B = \{0,2,6,8,10\}$ . Find:

(i)  $A \cap B$

Ans -  $\{2,3,4,5,6,7,8,10\}$

(ii)  $A \cup B$

Ans -  $A \cup B = \{0, 2, 3, 4, 5, 6, 7, 8, 10\}$

Q8. What do you understand about skewness in data?

**Ans - Skewness in data refers to the asymmetry of the distribution. It indicates whether the data is concentrated more on one side of the mean than the other. Positive skewness means a longer tail on the right side, negative skewness means a longer tail on the left side, and zero skewness indicates a symmetrical distribution.**

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

**Ans - Mean < Median < Mode**

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

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

Ans -

**The formula for calculating the sample mean is:**

$$\text{Sample Mean } (\bar{x}) = \Sigma x / n$$

**where:**

$\bar{x}$  = Sample mean

$\Sigma x$  = Sum of all values in the dataset

$n$  = Number of values in the sample

**Example calculation:**

**Let's calculate the sample mean for the following dataset:**

**[12, 15, 18, 20, 22, 25]**

**Step 1: Add up all the values in the dataset.**

$$\text{Sum of the values } (\Sigma x) = 12 + 15 + 18 + 20 + 22 + 25 = 112$$

**Step 2: Count the number of values in the dataset (n).**

$$\text{Number of values } (n) = 6$$

**Step 3: Apply the formula for the sample mean.**

**Sample Mean ( $\bar{x}$ ) =  $\Sigma x / n = 112 / 6 = 18.67$**

**The sample mean for the given dataset is approximately 18.67.**

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

**Ans - Mean = mode = median**

Q13. How is covariance different from correlation?

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

Note: Create your assignment in Jupyter notebook and upload it in GitHub & share that github repository link through your dashboard. Make sure the repository is public.

Data Science Masters