

Statistics Basics -2  
Assignment Questions – Solutions

Q1. What are the three measures of central tendency?

Ans : Measures of central tendency are Mean , Median & 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 represents the average value of database whereas median represents the middle value in a ascending order arranged dataset and in mode it defines by the most frequently occurring vale in dataset.

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 : Here mean = 176.98 , median = 177 , mode = 178

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 : Here mean = 177.38 and Variance = 3.0421 and Standard Deviation = 1.7438

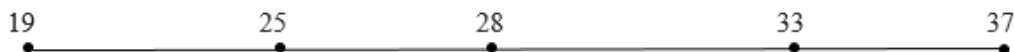
Q5. How are measures of dispersion such as range, variance, and standard deviation used to describe the spread of a dataset? Provide an example.

Ans :The range is the easiest dispersion of data or measure of variability. The range can measure by subtracting the lowest value from the massive Number. The wide range indicates high variability, and the small range specifies low variability in the distribution. To calculate a range, prepare all the values in ascending order, then subtract the lowest value from the highest value.

$$\text{Range} = \text{Highest\_value} - \text{Lowest\_value}$$

Eg .

Student_id	1	2	3	4	5
Marks	37	33	19	25	28



$$\begin{aligned}\text{Range} &= H - L \\ &= 37 - 19 \Rightarrow 18\end{aligned}$$

The range of marks is 18

Variance is a simple measure of dispersion. Variance measures how far each number in the dataset from the mean. To compute variance first, calculate the mean and squared deviations from a mean.

Population variance

$$\text{Variance} = \sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$$

Sample variance

Statistics Basics -2  
Assignment Questions – Solutions

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

### Standard Deviation

Standard deviation is a squared root of the variance to get original values. Low standard deviation indicates data points close to mean.

#### Population SD:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

#### Sample SD:

$$s = \sqrt{\frac{1}{N - 1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

Q6. What is a Venn diagram?

Ans : it is the visual representation of set , subset ,superset etc . It consists of one or more overlapping circles , where each circles represents a set and overlapping parts of the circles represent the elements that are common to those 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

Ans : A ∪ B = { 0,2,3,4,5,6,7,8,10} , A ∩ B = {2,6}

Q8. What do you understand about skewness in data?

Ans : Skewness is a measurement of the distortion of symmetrical distribution or asymmetry in a data set. It is the degree of asymmetry observed in a probability distribution. Distributions can exhibit right(positive) or left(negative) skewness to varying degrees. A normal distribution having bell curves exhibits zero skewness.

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

Ans : If the distribution of data is skewed to the right then mode is often less than median which is less than the mean i.e. mean is often greater than median.

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

Ans : Covariance reveals how two variables change together while correlation determines how closely two variables are related to each other. Correlation values are standardized but not of covariance values.

The covariance between X and Y is defined  
as  $\text{Cov}(X,Y) = E[(X - E(X))(Y - E(Y))] = E[XY] - (E(X))(E(Y))$ .

Correlation  $\rho(X,Y) = \text{cov}(X,Y) / \sigma_X \cdot \sigma_Y$

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

Ans : sample mean =  $\sum x_i / n$

Sample mean of 60, 57, 109, 50 =  $(60+57+109+50)/4=69$

Statistics Basics -2  
Assignment Questions – Solutions

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

Ans : For a normal distribution , the three measures of tendency (mean , median , mode) are equal and all located at the peak of the bell-shaped curve that represents the normal distribution.

Q13. How is covariance different from correlation?

Ans : Covariance shows how two variables differ ,whereas correlation shows how two variables are related. We can tell that covariance indicates the direction of linear relationship between variables and correlation measures both the strength and direction of liner relationship between two variables. Covariance is a measure of correlation.

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

Ans : outliers are numbers in a database that are vastly larger and smaller than the other values in the set. Mean is one type of central tendency i.e. always affected by an outlier. However median and mode are less affected by outliers , since they are based on middle values and the most common value , respectively.

Outliers also affect dispersion (range, variance, standard deviation) in different ways. If there are outliers significantly larger or smaller than the other values , the range will increase , in case of variance and standard deviation they will increase because they are far from the mean.

Eg. : database : 10,15,20,25,30,35,40,45,50,200

Here mean = 40 , range = 190 and they are greatly affected by outlier value. Value of median is 32.5 & it is less affected by outlier.