

Q1 There are three main measures of central tendency: mode, median, mean.

Q2 The mean (average) of a data set is found by adding all numbers in the data set and then dividing by the number of values in the set. The median is the middle value when a data set is ordered from least to greatest. The mode is the number that occurs most often in a data set. Ex: retirement age of 11 people, in whole years: 54, 54, 54, 55, 56, 57 Mode: 54
median: 54.5 Mean: 55

Q3 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] mode: 178
mean: 177.01875 median: 179.5

Q4 [178, 177, 176, 177, 178.2, 178, 175, 179, 180, 175, 178.9, 176.2, 177, 172.5, 178, 176.5] std=2.17

Q5 Two data sets can have the same mean but they can be entirely different. Thus to describe data, one needs to know the extent of variability. This is given by the measures of dispersion. Range, interquartile range, and standard deviation are the three commonly used measures of dispersion.

1. Range: The range is the difference between the largest and the smallest observation in the data. The prime advantage of this measure of dispersion is that it is easy to calculate. On the other hand, it has a lot of disadvantages. 2. Variance: The variance of the data is given by measuring the distance of the observed values from the mean of the distribution. Here we are not concerned with the sign of the distance of the point, we are more interested in the magnitude. 3. Standard Deviation: In the calculation of variance, notice that the units of the variance and the unit of the observations are not the same. So, to remove this problem, we define standard deviation. It is denoted as σ .

Q6 A Venn diagram is an illustration that uses circles to show the relationships among things or finite groups of things. Circles that overlap have a commonality while circles that do not overlap do not share those traits. Venn diagrams help to visually represent the similarities and differences between two concepts.

Q7 sets $A = \{2, 3, 4, 5, 6, 7\}$ & $B = \{0, 2, 6, 8, 10\}$. $A \cap B = \{2, 6\}$ $A \cup B = \{0, 2, 3, 4, 5, 6, 7, 8, 10\}$

Q8 Skewness is a measurement of the distortion of symmetrical distribution or asymmetry in a data set. Skewness is demonstrated on a bell curve when data points are not distributed symmetrically to the left and right sides of the median on a bell curve.

Q9

One of the basic tenets of statistics that every student learns in about the second week of intro stats is that in a skewed distribution, the mean is closer to the tail in a skewed

distribution. So in a right skewed distribution (the tail points right on the number line), the mean is higher than the median.

For a right skewed distribution, the mean is typically greater than the median. Also notice that the tail of the distribution on the right hand (positive) side is longer than on the left hand side. From the box and whisker diagram we can also see that the median is closer to the first quartile than the third quartile.

Q10 Covariance is a measure to indicate the extent to which two random variables change in tandem. Correlation is a measure used to represent how strongly two random variables are related to each other. Covariance is nothing but a measure of correlation. Correlation refers to the scaled form of covariance.

Both covariance and correlation measure the relationship and the dependency between two variables. Covariance indicates the direction of the linear relationship between variables. Correlation measures both the strength and direction of the linear relationship between two variables. Correlation values are standardized.

Q11 The general formula for calculating the sample mean is given by $\bar{x} = (\sum x_i) / n$. Here, \bar{x} represents the sample mean, x_i refers all X sample values and n stands for the number of sample terms in the data set.

Ex:- $I=[1,2,3,4,5,6,7,8,9,10]$ mean= $55/10=5.5$

Q12

When you have a normally distributed sample you can legitimately use both the mean or the median as your measure of central tendency. In fact, in any symmetrical distribution the mean, median and mode are equal. The three measures: mean, median, and mode under a normal distribution are all the same. Also, in a positively skewed distribution, the mean is the greatest number, as to the median and mode.

Q13 Covariance is when two variables vary with each other, whereas Correlation is when the change in one variable results in the change in another variable.

Q14 Outliers are numbers in a data set that are vastly larger or smaller than the other values in the set. Mean, median and mode are measures of central tendency. Mean is the only measure of central tendency that is always affected by an outlier. Mean, the average, is the most popular measure of central tendency. Outliers are extreme, or atypical data value(s) that are notably different from the rest of the data. It is important to detect outliers within a distribution, because they can alter the results of the data analysis. The mean is more sensitive to the existence of outliers than the median or mode. EX:- Consider the initial retirement age dataset again, with one difference; the last observation of 60 years has been replaced with a retirement age of 81 years. This value is much higher than the other values, and could be considered an outlier. However, it has not changed the middle of the

distribution, and therefore the median value is still 57 years.

54, 54, 54, 55, 56, 57, 57, 58, 58, 60, 81 As the all values are included in the calculation of the mean, the outlier will influence the mean value.

$(54+54+54+55+56+57+57+58+58+60+81 = 644)$, divided by 11 = 58.5 years

In []:

In []: