Q1. What are the three measures of central tendency?

1)mean

2)median

3)mode

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

The mode is the most frequent value. The median is the middle number in an ordered data set. The mean is the sum of all values divided by the total number of values.

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]

1)mean = 177.018

2)median = 177

3)mode = 177

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]

In [2]:

```python
import numpy as np
x  = [178,177,176,177,178.2,178,175,179,180,175,178.9,176.2,177,172.5,178,176.5]
y = np.std(x)
print(y)
```

1.7885814036548633

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

Measures of dispersion describe the spread of the data. They include the range, interquartile range, standard deviation and variance. The range is given as the smallest and largest observations

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.

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

Q6. What is a Venn diagram?

A Venn diagram illustrates the relationships between two or more data 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

(i) A U B = (0,2,3,4,5,6,7,8,10) (ii) A intersection B = (2,6)

Q8. What do you understand about skewness in data?

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. If the bell curve is shifted to the left or the right, it is said to be skewed.

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

mean>median>mode

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

Covariance is an indicator of the extent to which 2 random variables are dependent on each other. A higher number denotes higher dependency. Correlation is a statistical measure that indicates how strongly two variables are related

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

x̄ = (Σ xi) / n

ex : (100, 1, 25, 49, 0, 100, 49) x̄ = 324/7 = 46.3

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

normal distribution has a graph that is perfectly symmetric about a vertical line through its peak. Therefore, all measures of central tendency (most commonly, the mean, median, and mode) give the same answer.

mean = median = mode

Q13. How is covariance different from correlation?

Covariance is an indicator of the extent to which 2 random variables are dependent on each other. A higher number denotes higher dependency. Correlation is a statistical measure that indicates how strongly two variables are related

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

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.

Outliers Measures of central tendency and dispersion can give misleading impressions of a data set if the set contains one or more outliers. An outlier is a value that is much greater than or much less than most of the other values in a data set.