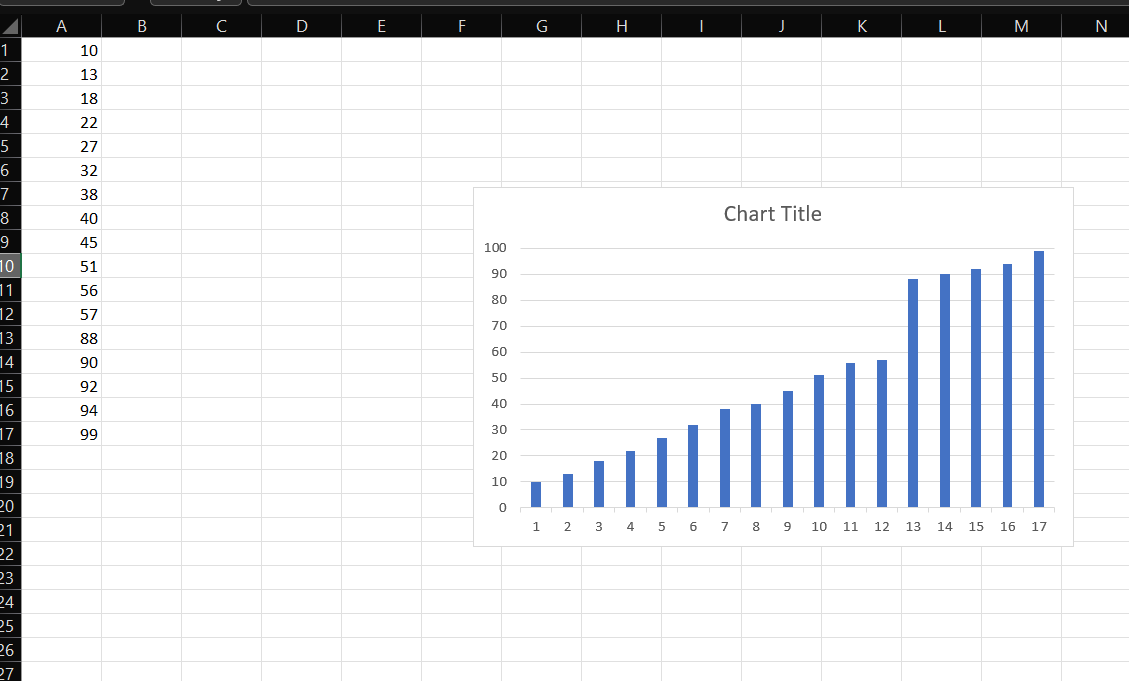
STATISTICS ASSIGNMENT

Que 1) Plot a histogram,

10, 13, 18, 22, 27, 32, 38, 40, 45, 51, 56, 57, 88, 90, 92, 94, 99



Que 2) In a quant test of the CAT Exam, the population standard deviation is known to be 100. A sample of 25 tests taken has a mean of 520. Construct an 80% CI about the mean.

Ans-

Determine the critical value for an 80% confidence level. Since the sample size is large (n = 25) and the population standard deviation is known, you can use the Z-distribution. Look up the critical value for an 80% confidence level in the Z-table or use a Z-table calculator. The critical value for an 80% confidence level is approximately 1.282.

Calculate the margin of error (ME) using the formula:

ME = Critical Value \* (Population Standard Deviation / Square Root of Sample Size)

ME = 1.282 \* (100 / √25) = 1.282 \* (100 / 5) = 1.282 \* 20 = 25.64

So, the margin of error is approximately 25.64.

Calculate the lower and upper bounds of the confidence interval:

Lower Bound = 520 - 25.64 = 494.36

Upper Bound = 520 + 25.64 = 545.64

Therefore, the 80% confidence interval about the mean is approximately 494.36 to 545.64.

Que 3) A car believes that the percentage of citizens in city ABC that owns a vehicle is 60% or less. A sales manager disagrees with this. He conducted a hypothesis testing surveying 250 residents & found that 170 residents responded yes to owning a vehicle.

1. State the null & alternate hypothesis.
2. At a 10% significance level, is there enough evidence to support the idea that vehicle owner in ABC city is 60% or less.

Ans-

In this scenario, we can state the null and alternative hypotheses as follows:

Null Hypothesis (H₀): The percentage of citizens in city ABC that owns a vehicle is 60% or less.

Alternative Hypothesis (H₁): The percentage of citizens in city ABC that owns a vehicle is greater than 60%.

To determine if there is enough evidence to support the idea that the vehicle ownership in ABC city is 60% or less, we will conduct a hypothesis test at a 10% significance level.

In hypothesis testing, we use a test statistic and compare it to a critical value or calculate a p-value to make a decision. Since we have sample data, we can use a z-test for proportions in this case.

Here's how we can perform the hypothesis test:

Determine the significance level (α): In this case, the significance level is given as 10% or 0.10.

Set up the critical region: Since the alternative hypothesis is one-sided (greater than), we will use a right-tailed test. At a 10% significance level, the critical z-value is approximately 1.28 (obtained from the standard normal distribution table).

Calculate the test statistic (z-score):

z = (p̂ - p₀) / √((p₀ \* (1 - p₀)) / n)

z = (0.68 - 0.60) / √((0.60 \* (1 - 0.60)) / 250) = 2.02

Compare the test statistic with the critical value: The calculated z-score of 2.02 is greater than the critical value of 1.28.

Make a decision: Since the test statistic falls in the critical region, we reject the null hypothesis.

State the conclusion: At a 10% significance level, there is enough evidence to support the idea that the percentage of citizens in city ABC who own a vehicle is greater than 60%.

In summary, based on the provided data and the hypothesis test at a 10% significance level, there is enough evidence to support the notion that the vehicle ownership percentage in ABC city exceeds 60%.

Que 4) What is the value of the 99 percentile?

2,2,3,4,5,5,5,6,7,8,8,8,8,8,9,9,10,11,11,12

Ans-

To find the 99th percentile value for a given dataset, we need to find the value below which 99% of the data falls. Here's how it's calculated:

Sort the dataset in ascending order:

2, 2, 3, 4, 5, 5, 5, 6, 7, 8, 8, 8, 8, 8, 9, 9, 10, 11, 11, 12

Calculate the index corresponding to the 99th percentile using the formula:

Index = (Percentile / 100) \* (n + 1)

Index = (99 / 100) \* (20 + 1) = 0.99 \* 21 = 20.79

Identify the values at the index and the next highest index. Since the index is not an integer, you will take the floor and ceil values. In this case, the floor index is 20, and the ceil index is 21.

The 99th percentile value will be the average of the values at the floor and ceil indices:

99th Percentile = (Value at floor index + Value at ceil index) / 2

In this case, the value at the floor index is 11, and the value at the ceil index is 12.

99th Percentile = (11 + 12) / 2 = 11.5

Therefore, the value of the 99th percentile in the given dataset is 11.5.

Que 5) In left & right-skewed data, what is the relationship between mean, median & mode?

Draw the graph to represent the same.

Ans-

In left-skewed data, the relationship between the mean, median, and mode can be described as follows:

Mean: The mean is typically less than the median in left-skewed data. This is because the tail of the distribution is stretched towards the left, pulling the mean in that direction.

Median: The median is generally greater than the mean in left-skewed data. This is because the median represents the middle value of the dataset, and in left-skewed distributions, the lower values are more frequent and tend to pull the median towards the left.

Mode: The mode is the value or values that appear most frequently in the dataset. In left-skewed data, the mode is generally greater than the mean and median. This is because the peak of the distribution is shifted towards the right, resulting in a higher frequency of higher values.

Graphically, a left-skewed distribution would have a longer tail on the left side, indicating a higher frequency of lower values. The mean would be pulled towards the tail, while the median would be closer to the center but still shifted towards the left. The mode would typically be the highest point on the distribution, representing the most frequent value(s).

Here is an example graph depicting a left-skewed distribution:

\*

\*

\*

\*

\*

-----------------------------

In right-skewed data, the relationship between the mean, median, and mode is reversed:

Mean: The mean is typically greater than the median in right-skewed data. This is because the tail of the distribution is stretched towards the right, pulling the mean in that direction.

Median: The median is generally less than the mean in right-skewed data. This is because the median represents the middle value of the dataset, and in right-skewed distributions, the higher values are more frequent and tend to pull the median towards the right.

Mode: The mode is the value or values that appear most frequently in the dataset. In right-skewed data, the mode is generally less than the mean and median. This is because the peak of the distribution is shifted towards the left, resulting in a higher frequency of lower values.

Graphically, a right-skewed distribution would have a longer tail on the right side, indicating a higher frequency of higher values. The mean would be pulled towards the tail, while the median would be closer to the center but still shifted towards the right. The mode would typically be the highest point on the distribution, representing the most frequent value(s).

Here is an example graph depicting a right-skewed distribution:

-----------------------------

\*

\*

\*

\*

\*

It's important to note that the specific positions of the mean, median, and mode can vary depending on the shape and characteristics of the dataset, but the general relationships described above hold true in left-skewed and right-skewed distributions.