**Statistics Assignment 1**

**Q1. Plot a histogram,**

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

Ans:

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

num\_bins = 5

bin\_size = 20

* (0-20): 10, 13, 18
* (20-40): 22, 27, 32, 38
* (40-60): 40, 45, 51, 56, 57
* (60-80): N/A
* (80-100): 88, 90, 92, 94, 99

A picture containing text, screenshot, diagram, rectangle

Description automatically generated

**Q2. 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: To construct an 80% confidence interval (CI) about the mean, we can use the formula:

CI = mean ± (critical value \* standard deviation / sqrt (sample size))

Given: Population standard deviation (σ) = 100

Sample mean (x̄) = 520

Sample size (n) = 25

First, we need to find the critical value corresponding to an 80% confidence level. Since the sample size is small (n < 30), we can use a t-distribution instead of a normal distribution.

Using a t-distribution table or a statistical software, we find that the critical value for an 80% confidence level and a sample size of 25 is approximately 1.319.

Now, let's calculate the confidence interval:

CI = 520 ± (1.319 \* 100 / sqrt (25))

= 520 ± (1.319 \* 100 / 5)

= 520 ± 26.38

≈ (493.62, 546.38)

**Q3. 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: Null Hypothesis (H0): The percentage of citizens in city ABC who own a vehicle is 60% or less. Alternative Hypothesis (H1): The percentage of citizens in city ABC who own a vehicle is greater than 60%.

To test these hypotheses, we can perform a one-sample proportion test. We will compare the observed proportion of vehicle owners in the sample to the hypothesized proportion of 60%.

Let's calculate the test statistic and p-value using the provided information:

Observed proportion of vehicle owners: p̂ = 170/250 = 0.68

Expected proportion (under the null hypothesis): p0 = 0.60

Standard error: SE = sqrt(p0 \* (1 - p0) / n) = sqrt(0.60 \* 0.40 / 250) ≈ 0.0346

Test statistic (z-score): z = (p̂ - p0) / SE = (0.68 - 0.60) / 0.0346 ≈ 2.19

At a 10% significance level (α = 0.10), we need to compare the test statistic to the critical value. Since we are testing the alternative hypothesis that the percentage of vehicle owners is greater than 60%, we will perform a one-tailed test.

Using a standard normal distribution table or statistical software, we find the critical z-value for a one-tailed test at a significance level of 0.10 to be approximately 1.28.

Since the calculated test statistic (z = 2.19) is greater than the critical value (1.28), we can reject the null hypothesis.

**Q4. What is the value of the 99 percentiles?**

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

Ans: To find the value of the 99th percentile in the given dataset, we need to determine the value below which 99% of the data falls.

First, let's arrange 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

The percentile indicates the percentage of values that are less than or equal to a certain value. To calculate the 99th percentile, we need to find the position of the value that corresponds to the 99th percentile.

The formula to calculate the position (index) of a percentile is:

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

where n is the total number of data points.

In this case, we have 20 data points (n = 20), and we want to find the value corresponding to the 99th percentile.

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

Since the index is not an integer, we need to interpolate to find the value at the 99th percentile.

Interpolation involves taking the weighted average of the values surrounding the index.

The value at the 20th index is 11, and the value at the 21st index is 12.

To interpolate, we can use the formula:

Value = (1 - decimal) \* value\_at\_lower\_index + decimal \* value\_at\_higher\_index

In this case, the decimal is 0.79 (the decimal part of the index).

Value = (1 - 0.79) \* 11 + 0.79 \* 12

= 0.21 \* 11 + 0.79 \* 12

≈ 2.31 + 9.48

≈ 11.79

Therefore, the value of the 99th percentile in the given dataset is approximately 11.79.

Q5. In left & right-skewed data, what is the relationship between mean, median & mode?

Draw the graph to represent the same.

Ans:

* **In left-skewed (negatively skewed) data:**
* The mean is generally less than the median.
* The median is usually closer to the right tail of the distribution.
* The mode is typically the highest point of the distribution.
* **In right-skewed (positively skewed) data:**
* The mean is generally greater than the median.
* The median is usually closer to the left tail of the distribution.
* The mode is typically the highest point of the distribution.
* **Graphical representation:**

