**Statistics Assignment**

**Que 1) Plot a histogram,**

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

import matplotlib.pyplot as plt

# Given data

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

# Plot histogram

plt.hist(data, bins=10, color='skyblue', edgecolor='black')

# Add labels and title

plt.xlabel('Value')

plt.ylabel('Frequency')

plt.title('Histogram of Data')

# Show plot

plt.show()

**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.**

To construct a confidence interval (CI) for the population mean when the population standard deviation is known, you can use the z-score formula. The formula for calculating the CI is:

\[ \text{CI} = \bar{x} \pm z \left( \frac{\sigma}{\sqrt{n}} \right) \]

Where:

- \(\bar{x}\) is the sample mean,

- \(z\) is the z-score corresponding to the desired confidence level,

- \(\sigma\) is the population standard deviation, and

- \(n\) is the sample size.

For an 80% confidence interval, we need to find the z-score corresponding to the middle 80% of the standard normal distribution. This corresponds to the z-scores at the 10th and 90th percentiles, which are -1.28 and 1.28, respectively.

Given:

- Sample mean (\(\bar{x}\)) = 520

- Population standard deviation (\(\sigma\)) = 100

- Sample size (\(n\)) = 25

Let's calculate the confidence interval:

\[ \text{CI} = 520 \pm 1.28 \left( \frac{100}{\sqrt{25}} \right) \]

\[ \text{CI} = 520 \pm 1.28 \left( \frac{100}{5} \right) \]

\[ \text{CI} = 520 \pm 1.28 \times 20 \]

\[ \text{CI} = 520 \pm 25.6 \]

Now, we can construct the confidence interval:

Lower Limit = 520 - 25.6 = 494.4

Upper Limit = 520 + 25.6 = 545.6

So, the 80% confidence interval about the mean is (494.4, 545.6).

**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.**

a. Null Hypothesis (\(H\_0\)): The percentage of citizens in city ABC that owns a vehicle is 60% or less.

Alternate Hypothesis (\(H\_a\)): The percentage of citizens in city ABC that owns a vehicle is more than 60%.

b. To determine whether there is enough evidence to support the idea that the vehicle ownership in ABC city is 60% or less, we can conduct a one-sample proportion z-test.

Given:

- Sample size (\(n\)) = 250

- Number of residents responded yes (\(x\)) = 170

- Expected proportion (from the null hypothesis) (\(p\_0\)) = 0.60

- Significance level (\(\alpha\)) = 0.10

We can calculate the test statistic (\(z\)) using the formula:

\[ z = \frac{\hat{p} - p\_0}{\sqrt{\frac{p\_0 \cdot (1 - p\_0)}{n}}} \]

where \(\hat{p}\) is the sample proportion.

\[ \hat{p} = \frac{x}{n} = \frac{170}{250} = 0.68 \]

\[ z = \frac{0.68 - 0.60}{\sqrt{\frac{0.60 \cdot (1 - 0.60)}{250}}} \]

\[ z = \frac{0.68 - 0.60}{\sqrt{\frac{0.60 \cdot 0.40}{250}}} \]

\[ z = \frac{0.08}{\sqrt{\frac{0.24}{250}}} \]

\[ z = \frac{0.08}{\sqrt{0.00096}} \]

\[ z \approx \frac{0.08}{0.0310} \]

\[ z \approx 2.581 \]

Now, we compare the calculated \(z\) value with the critical value from the standard normal distribution table at \(\alpha = 0.10\) significance level. At \(\alpha = 0.10\), the critical value is approximately \(z = 1.28\) (for a one-tailed test).

Since \(z = 2.581 > 1.28\), we reject the null hypothesis (\(H\_0\)).

Conclusion: At a 10% significance level, there is enough evidence to support the idea that the vehicle ownership in ABC city is more than 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**

To find the 99th percentile of a dataset, we need to determine the value below which 99% of the data falls. We can achieve this by first arranging the data in ascending order and then identifying the position of the 99th percentile.

Given dataset:

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

1. Arrange the data in ascending order:

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

2. Calculate the position of the 99th percentile:

\( n = 20 \) (total number of data points)

\( 99\% \) of \( n \) is \( 0.99 \times 20 = 19.8 \)

Since the position is not a whole number, we take the value at the next highest position. So, the 99th percentile is at the 20th position.

3. Find the value at the 20th position, which corresponds to the 99th percentile:

The value at the 20th position in the sorted dataset is 12.

Therefore, the 99th percentile of the given dataset is 12.

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

**Draw the graph to represent the same.**

In left-skewed (negatively skewed) data:

- Mean < Median < Mode

In right-skewed (positively skewed) data:

- Mode < Median < Mean

To represent this relationship graphically, let's plot histograms for left-skewed and right-skewed distributions and illustrate the positions of mean, median, and mode.

First, let's create some example data for both distributions and then plot histograms:

```python

import numpy as np

import matplotlib.pyplot as plt

# Generate left-skewed data

left\_skewed\_data = np.random.exponential(scale=2, size=1000)

# Generate right-skewed data

right\_skewed\_data = 10 + np.random.exponential(scale=2, size=1000)

# Plot histograms

plt.figure(figsize=(12, 6))

# Left-skewed histogram

plt.subplot(1, 2, 1)

plt.hist(left\_skewed\_data, bins=30, color='skyblue', edgecolor='black')

plt.axvline(np.mean(left\_skewed\_data), color='red', linestyle='dashed', linewidth=1, label='Mean')

plt.axvline(np.median(left\_skewed\_data), color='green', linestyle='dashed', linewidth=1, label='Median')

plt.legend()

plt.title('Left-Skewed Data')

# Right-skewed histogram

plt.subplot(1, 2, 2)

plt.hist(right\_skewed\_data, bins=30, color='salmon', edgecolor='black')

plt.axvline(np.mean(right\_skewed\_data), color='red', linestyle='dashed', linewidth=1, label='Mean')

plt.axvline(np.median(right\_skewed\_data), color='green', linestyle='dashed', linewidth=1, label='Median')

plt.legend()

plt.title('Right-Skewed Data')

plt.tight\_layout()

plt.show()

```

In the histograms, the dashed red line represents the mean, the dashed green line represents the median, and the tallest bar represents the mode.

- In the left-skewed histogram, the mean < median < mode.

- In the right-skewed histogram, the mode < median < mean.

These relationships visually illustrate the positions of mean, median, and mode in left-skewed and right-skewed distributions.