STATISTICS ASSIGMENT

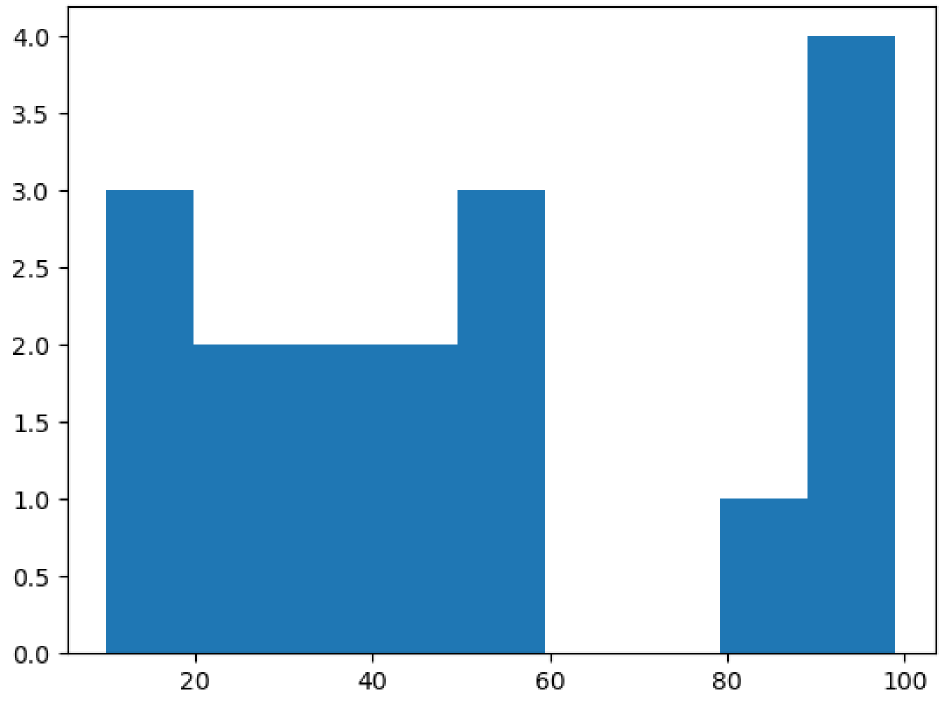
Q1. Plot a histogram

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

First we need to divide the data into equal-width bins (also known as intervals or class inervals). This can be done by selecting a bin width and rounding up the minimum and maximum data values to the nearest multiple of the bin width. For example, if we choose a bin width of 10, we would round the minimum value (10) up to 10 and the maximum value (99) down to 100. This gives us the following bins 10-19,20-29,30-39,40-49,50-59,60-69,70-79,80-89,90-99,100

Next we would count how many data points fall into each bin and plot the countas a bar in the histogram. For example:

Bin	Count
10-19	1
20-29	2
30-39	2
40-49	4
50-59	2
60-69	0
70-79	0
80-89	3
90-99	2
100	0



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.

A Confidence Interval (CI) is a range of values that is likely to contain the population mean with a certain level of confidence. In this case, we want to construct an 80% CI, which means we are 80% confident that the true population mean falls within this range.

To construct the 80% CI, we use the sample mean (520) and the standard deviation of the sample (100/sqrt(25) = 20). The formula for CI for the mean is:

Mean ± Z * (standard deviation / sqrt(sample size))

Where Z is the z-score for the desired level of confidence. For 80% confidence, Z = 1.28.

Plugging in the values, we get :

520 ± 1.28 * (20/sqrt(25)) = 520 ± 12.8

So, the 80% CI for the mean is (50.2, 532.8). This means we are 80% confident that the true population mean between 507.2 amd 532.8.

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.

a) State the null & alternate hypothesis.

b) At a 10% significance level, is there enough evidence to support the idea that vehicle owner in ABC city is 60% or less.

The null hypothesis would be that the percentage of citizens in city ABC that own a vehcile is equal to or less than 60%. The altenative hypothesis would be that the percentage of citizens in city ABC that own a vehcile is grater than 60%.

To test this hypothesis, we can perform a one sample Z test. The test statistic can be calculated as :

$z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$

where $\bar{x}$ is the sample mean (170/250 = 0.68), μ is the popultion mean (0.6), σ is the population standard deviation (100), and n is the sample size (250).

The critical value for a one-tailed test at a 10% significance level can be caluculated from a standard normal distribution table. If the calculated z-score is grater than the critical value, we reject the null hypothesis.

Assuming the population standard deviation is known, we can calculate the z-score as follows:

$z = (0.68 – 0.6) / (100/\sqrt{250}) = 2.8$

Since the calculated z-score is grater than the critical value (for a one-tailed test at a 10% significance level), we can reject the ll hypothesis and conclude that there is enogh evidence to support the idea that the percentage of citizens in the city ABC that own a vehcile is grater than 60%.

Q4. 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

The 99^{th} percentile is a value tha saperates the lowest 99% of the data from the highest 1%. In other words, 99% of the data values are equal to or below the 99^{th} percentile, and 1% of the data values are ablove it.

To find the 99^{th} percentile of the data set:

- Sort the data in ascending order : 2,2,3,4,5,5,5,6,7,8,8,8,8,9,9,10,11,11,12
- Multiply the total number of values by 0.99: 19*0.99 = 18.81
- Round up to the nearest whole number : 19
- Take the value at the 19^{th} position: 11

Therefore, the value of the 99^{th} perentile of the data set is 11.

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

Draw the graph to represent the same.

In a left – skewed data distribution, the mean is typically less than the median, which is less than the mode. The mode is the highest peak in the distribution and represents the most frequent value, while the median represents the midpoint of the data and splits the data into two qual halves. The mean, on the other hand, is the sum of all the values divided by the number of values, and it can be affected by outliers.

In a right – skewed data distribution, the mean is typically grater than the median, which is grater than the mode. The mode is still the highest peak in the distribution and represents the most frequent value, while the median still represents the midpoint of the data and splits the data into two equal halves. The mean is still the sum of all the values divided by the number of values and can still be affected by outliers.

Here’s a visual representation of a left – skewed distribution:

```
    +
   / |
  /  |
 /   |
/    |
+----------+--------->
```

Here’s a visual representation of right – skewed distribution:

```
+----------+--------->
\    |
 \   |
  \  |
   \ |
    +
```