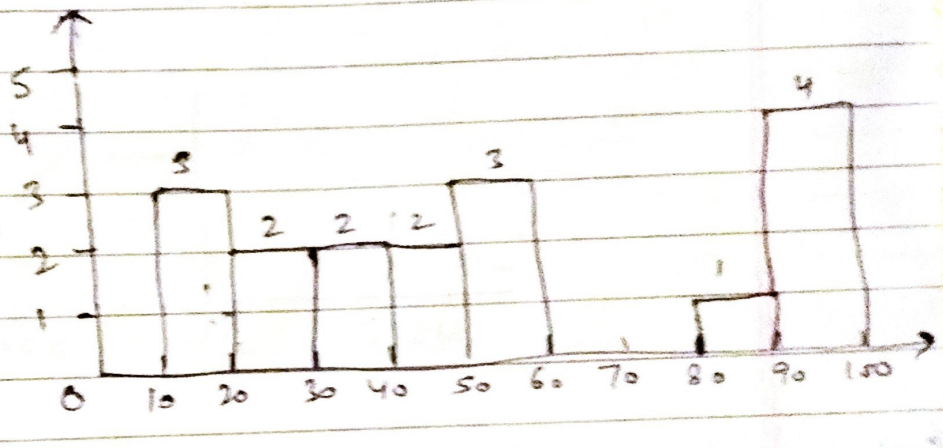
STATISTICS ASSIGNMENT

Ques1: Plot a histogram :-

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

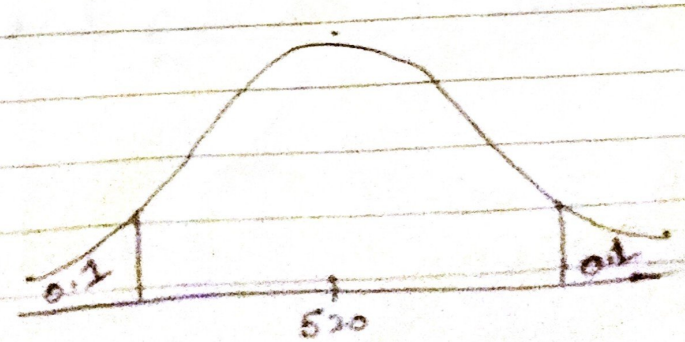
Ans Assuming Bin size as 10, Data is sorted in ascending order

Grouping	Data
0-10	0
10-20	3
20-30	2
30-40	2
40-50	2
50-60	3
60-70	0
70-80	0
80-90	1
90-100	4



Ques2: In a quant test of 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 $\sigma = 100$, $n = 25$, $\bar{x} = 520$, $CI = 0.8$, $\alpha = 1 - 0.8 = 0.2$



Lower fence = $520 - Z_{\frac{0.1}{2}} \frac{100}{\sqrt{25}}$

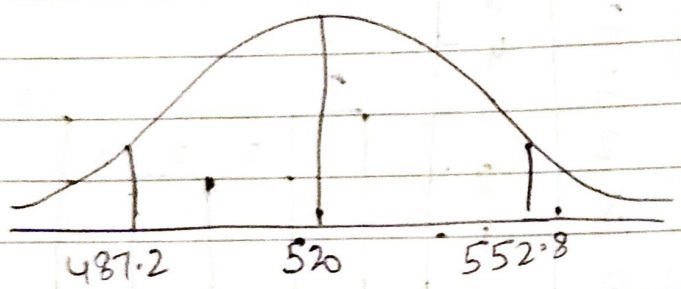
$= 520 - Z_{0.05} \times 20$

$= 520 - 1.64 \times 20$

$= 520 - 32.8 = 487.2$

Higher fence = $520 + 32.8$

$= 552.8$



Ques3 A Car belives 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.

Ans

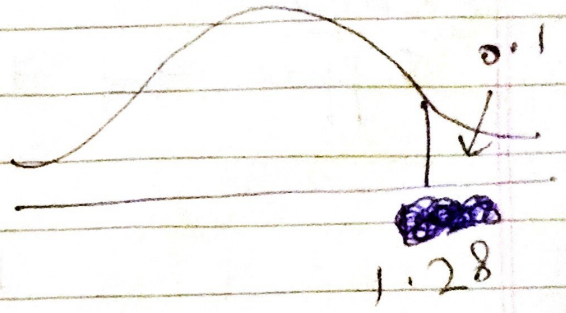
(a) H_0 : $P_0 \leq 60\%$

$P_0 \geq 60\%$

(b) $p = 0.6$ $\alpha = 0.1$

$q = 1 - 0.6 = 0.4$

$\hat{p} = \frac{170}{250} = 0.68$



$$\text{Z-test} = \frac{\hat{p} - p^0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.68 - 0.6}{\sqrt{\frac{0.6 \times 0.4}{250}}} \approx 2.67$$

$2.67 > 1.28$ [Reject the null hypothesis]

Vehicle owners in ABC city are more than 60%.

Ques 4. What is the value of 99 percentile?

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

n = 20

99 Percentile $= \frac{99}{100} \times 20 = 19.8$

$\downarrow$

Avg of 19th & 20th

$= \frac{11+12}{2} = \frac{23}{2} =$ **11.5**

Ques 5 In left & right skewed data, what is the r'ship b/w mean, median and mode. Draw a graph for represent the same.

Ans In left skewed data :-

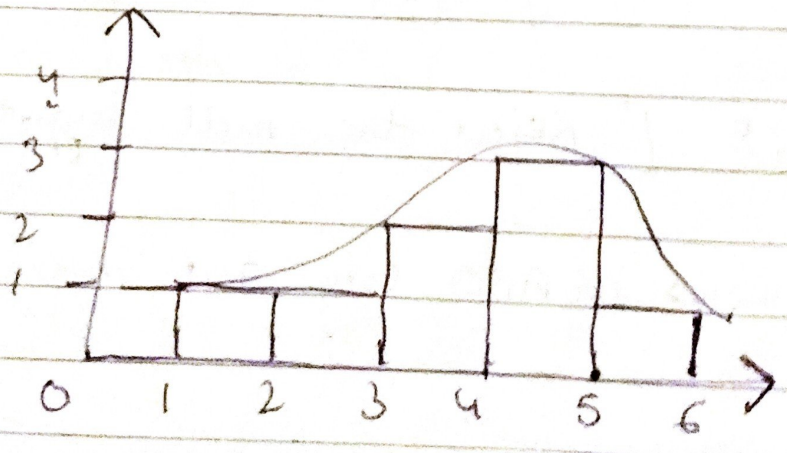
Mean < Median < Mode

Example :-

Suppose the data is → [illegible]

1, 2, 3, 3, 4, 4, 4, 5

Plotting a histogram with bin size as 1



1-2	1
2-3	1
3-4	2
4-5	3
5-6	1

Mean = $\frac{1+2+3+3+4+4+4+5}{8} = \frac{26}{8} = 3.25$

Median = 1, 2, 3, [3, 4] 4, 4, 5

$\frac{3+4}{2} = 3.5$

Mode = 4

So, $3.25 < 3.5 < 4$

Mean < Median < Mode

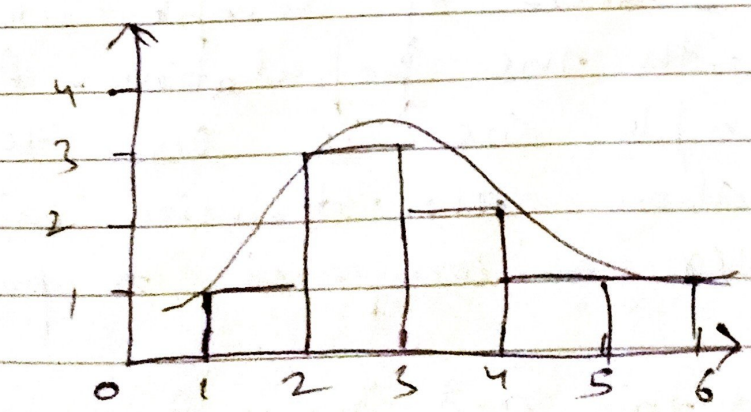
In Right skewed data :-

Mean > Median > Mode

Example :-

Suppose the data is 1, 2, 2, 2, 3, 3, 4, 5

Plotting a histogram with bin size as 1



1-2	1
2-3	3
3-4	2
4-5	1
5-6	1

Mean = $\frac{1+2+2+2+3+3+4+5}{8} = \frac{22}{8} = 2.75$

Median = 1, 2, 2, [2, 3] 3, 4, 5 = $\frac{2+3}{2} = 2.5$

Mode = 2

So, $2.75 > 2.5 > 2$

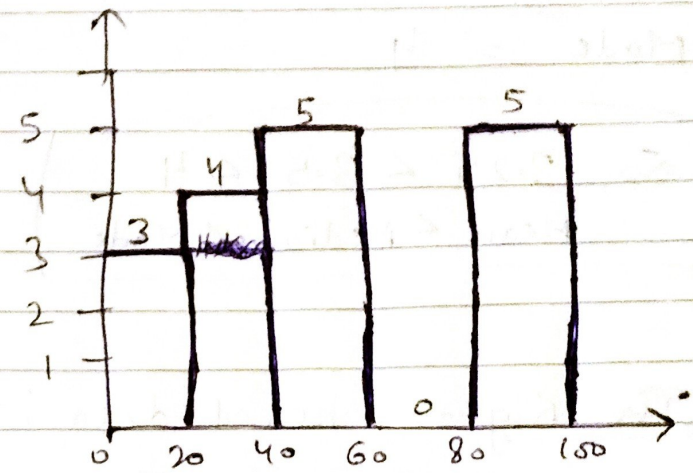
Mean > Median > Mode

Ques6 Plot a histogram

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

Bin Size = 20

Group	Data
0-20	3
20-40	4
40-60	5
60-80	0
80-100	5



Ques7 What is Bessel's correction?

Bessel's correction is the use of n-1 while calculating the standard deviation or variance of a sample. It is used to reduce the biasness of sample, when compared with the population. As if the sample size is less than the population the deviation will also be lesser as compared to population.

To correct this and bring it more equal to population the denominator i.e (n-1) is used which automatically increases the numerator resulting in a answer close to population.