**Que 1) Plot a histogram,**

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

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | AGE | FREQUENCY | | 0-20 | 3 | | 20-40 | 4 | | 40-60 | 5 | | 60-80 | 0 | | 80-100 | 5 | |  |
|  |  |
| BINS- 5  BIN SIZE – 20 |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**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.**



Given:

Sigma (population standard deviation) = 100

n(sample) = 25

xbar = 520 (sample mean)

To find: 80% of confidence interval about the mean

Formula:

Point estimator (+) or (-) Margin of Error

Point Estimator = xbar = 520

xbar (+) or (-) Zalpha/2  \* sigma/sq root of n

confidence interval 0f 80% can be taken as 0.8

Solution:

Significance value = 1- confidence interval = 1-0.8 = 0.2

Z0.2/2 = Z0.1

1-0.1=0.9

The Z value of 0.9 in Z-table is 1.29

So the range lies between -1.29 to 1.29

Lower fence= xbar – (Zalpha/2)value \* sigma/sq root of n

= 520-(1.29\*100/sq root of 25)

=520-(1.29\*20)

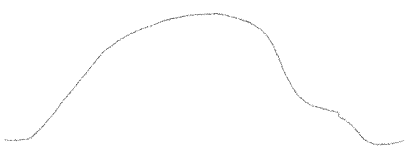
=520-25.8

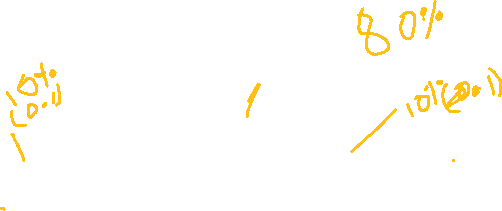
Lower fence value is 494.2

Higher fence= xbar + (Zalpha/2)value \* sigma/sq root of n

= 520+(1.29\*100/sq root of 25)

=520+25.8 = 545.8









Hence the confidence interval for the mean is between 494.8 to 545.8

**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.**

Given:

x=170 n=250

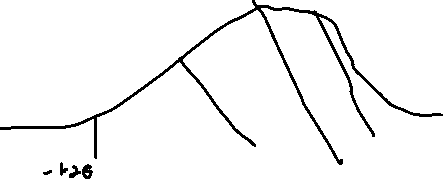
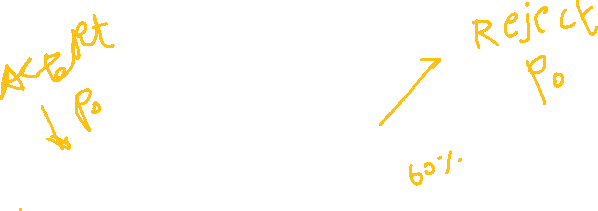
Significance level = 10%, hence alpha=0.1

Proportion= 60%

1. P0 <=60%

P1> 60%

We have an alpha value of 0.1 and hence we can draw the graph with this for 60%. Here the question is whether it is 60% or less







P^ = x/n = 170/250= 0.68

P0 = 0.6 (60% given)

Q0 = 1-P0 = 0.4

Alpha = 0.1 (10% significance level)

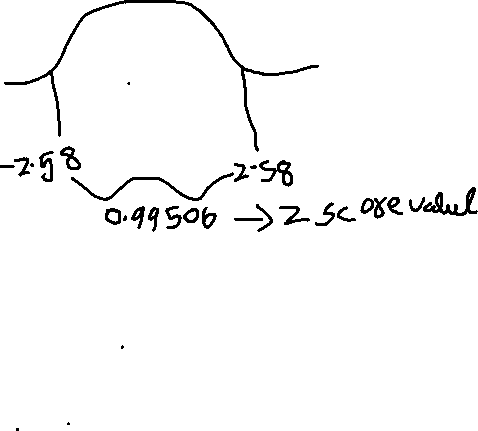
Z test with proportion:

= P^ - P0 / sq root of (P0 \*Q0)/n

= 0.65-0.6/sq root of (0.6) (0.4) / 250

= 2.5820

Finding P value



= 1- 0.99506

=0.00484

=0.00484 + 0.00484

P value = 0.00988

It is proven that the p-value (0.00988) is less than the significance level (0.1) and the z-score is 2.5820 which lies above -1.28, hence, the evidence is against the null hypothesis and there are more than 60% of vehicles in the city ABC.

**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 value for a specific percentile = percentile/100 \* n+ 1

n= number of data sets

= (99/100) \* 21 = 20.79 (the index value)

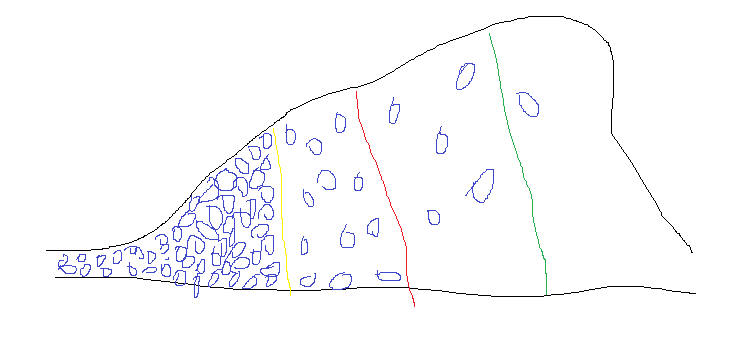
Since we do not have a 20.79 index we consider it as the 20th index and hence the value for the 99 percentile is 12.

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

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

Let’s have a look at the skews separately and discuss their relationship with mean, median, and mode with an example.

**Left skew:**



**Yellow- mean**

**Red- median**

**Green- mode**

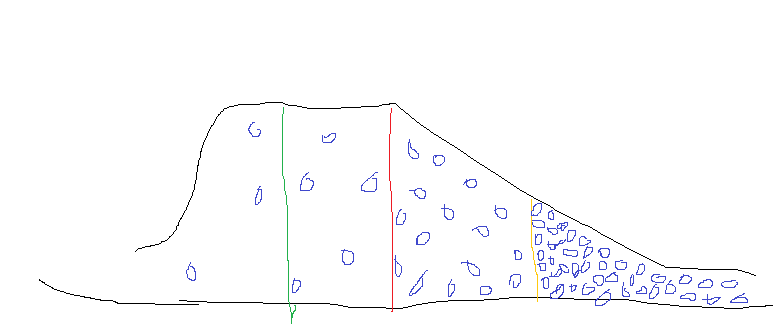


This is called left skew or negative skew, where most of the values et distributed in the range of the lowest limit to the median point. Let’s take the height of people from a particular country as an example and approach this concept, the above diagram states that the majority or maximum of the values in the dataset are spread in the lower limit and only a few values are above the upper limit, hence we can say that the above diagram the height Japanese people are spread over and the majority of their height lies below the mean value and somewhere near the right-hand side of the mean value. And close to 10% or fewer values i.e. from the mode (green colour) spread on the right side. Hence this is called left-skewed distribution means most of the values are on the left side of the minimal side.



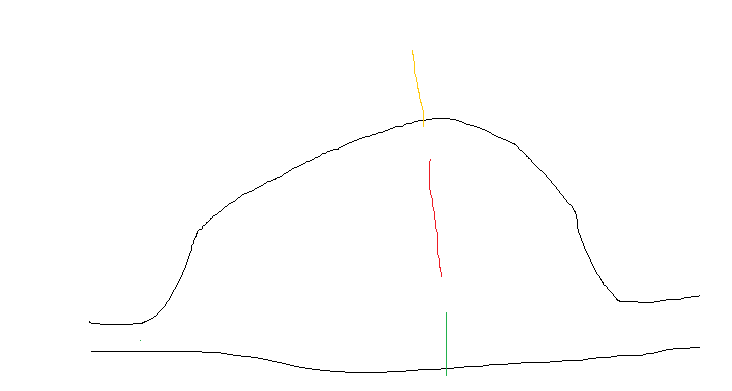
If the average height of the Japanese is 5.2 feet then most of the values fall within it(below the red mark) and there are minimal possibilities that the height of Japanese might fall from 5.2-5.5 feet (maybe for example) that gets rarely spread in the right side. Hence here the value of mode will be the highest (i.e. lesser number of tall people) and in terms of mean and median, the distribution range will be quite close with more number of people.

**Right skew:**

****

The right or the positive skew is opposite to that of the left skew concept, here the mean is the highest one followed by the median and then mode. We can also see more data got spread in the right-skewed region. Coming to our example we can say the Dutch people are the tallest race of humankind and hence we can see that more number of tall people lies on the right side of the distribution and it gradually getting decreased when it moves towards the left.

**Symmetrical or normal distribution**

****

In the symmetrical or normal distribution, the value of the mean, median and mode will almost be similar. It states that as per example there will be not much difference between the average height of a person when compared to the lower side that falls on the left side and the upper or above the average height and this scenario we can take for India.