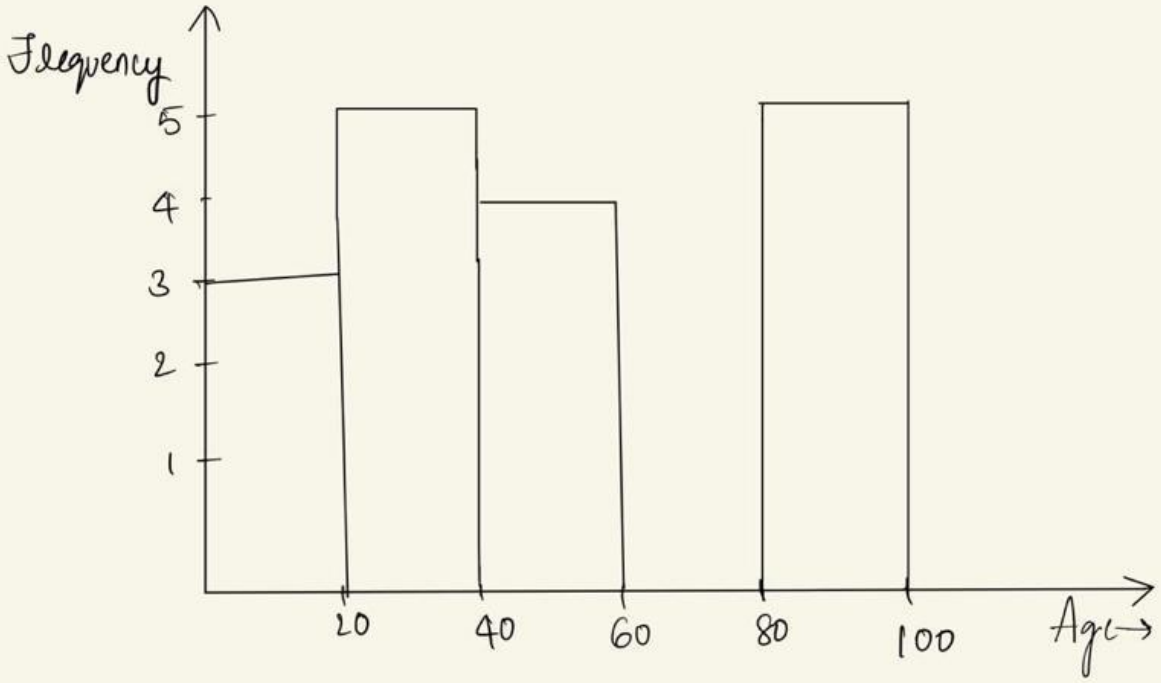
Assignment

Histogram of given variables.

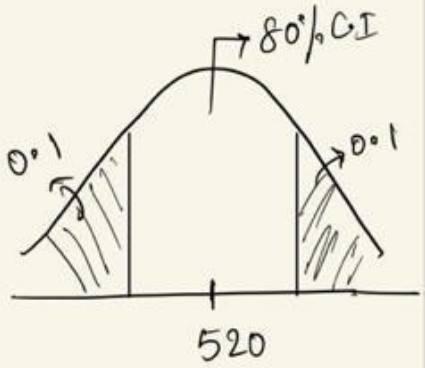
Q) {10, 13, 18, 22, 27, 32, 38, 40, 45, 51, 56, 57, 88, 90, 92, 94, 99} = Ages.

Bins = 5

Bins size = $\frac{100}{5}$ = 20



2) 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% confidence interval.



Given

$\sigma = 100$

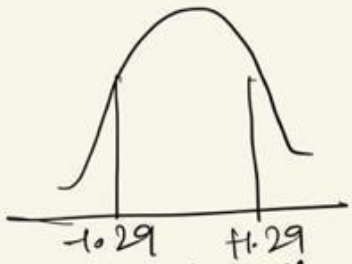
$n = 25$

$\bar{x} = 520$

Confidence interval = 80% = C.I

$\alpha = 1 - C.I = 1 - 0.8 = 0.2$

$\alpha/2 = 0.1$



Z test

Referring $Z_{(\alpha/2)} = Z_{0.1} = \underline{1.29}$ (referring the table)

Lower fence of C.I = point estimate − Margin of Error

Higher fence of C.I = point Estimate + Margin of Error.

where → margin of error = $Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$

↓

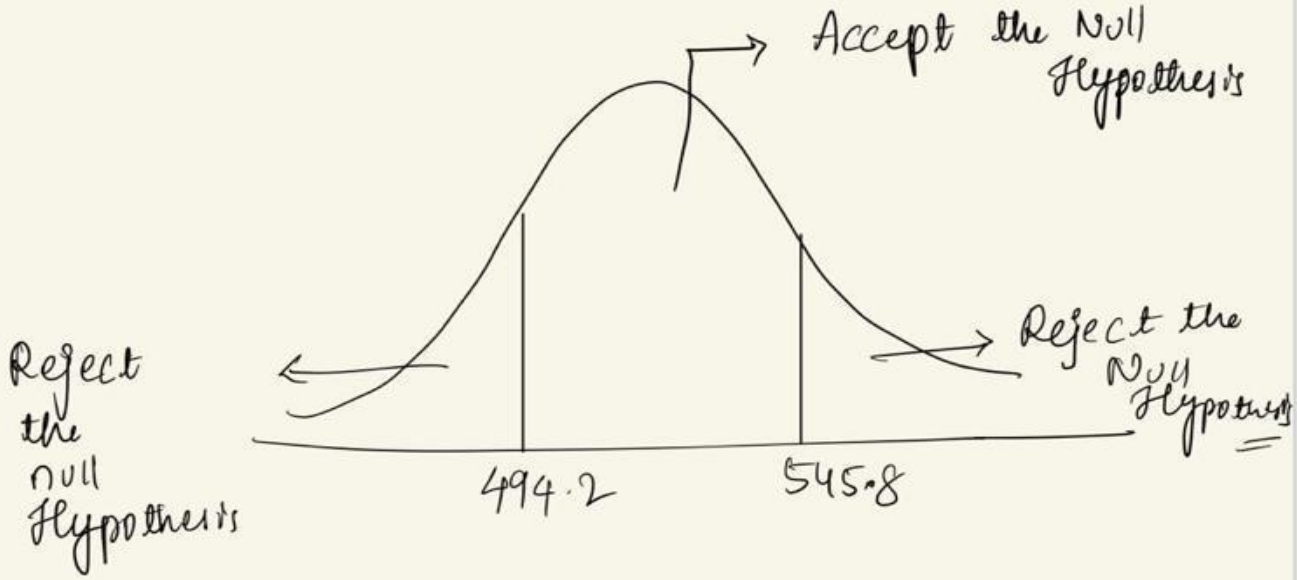
point Estimate = $\bar{x}$

$= Z_{0.1} \times \frac{100}{\sqrt{25}}$

$= 1.29 \times \frac{100}{5} = 25.8$

Lower fence of C.I = 520 − 25.8 = 494.2

Higher fence of C.I = 520 + 25.8 = 545.8



3) A car company 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 and 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:- Given info:- citizens owning $\Rightarrow$ 60% or less

$n = 250$

$x = 170$

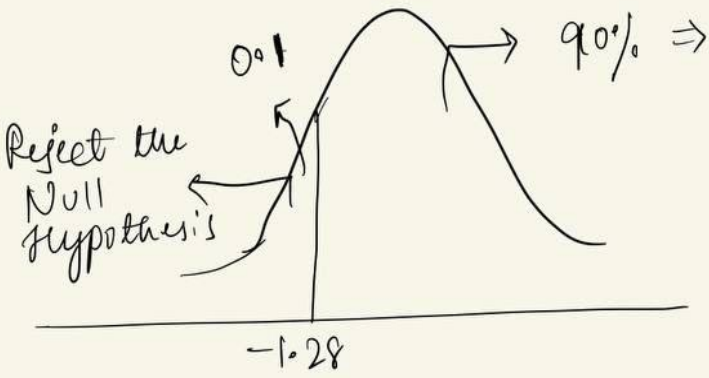
$\hat{P} = \frac{x}{n} = \frac{170}{250} = 0.68$

$\boxed{\alpha = 0.1}$

confidence interval $= 0.9$ i.e 90%

$H_0 \Rightarrow P_0 = \leq 60\%$ i.e citizens owning vehicles is 60% or less.

$H_1 \Rightarrow$ citizens owning vehicles is more than 60%



$Z = -1.28$ [referring the table]

$$Z_{test} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.68 - 0.6}{\sqrt{\frac{(0.6)(1-0.6)}{250}}}$$

$$= 2.5819$$

$1.28 < 2.5819 \Rightarrow$ Reject the Null Hypothesis

Conclusion:- Citizens owning vehicles is more than 60%

4) Value of the 99 percentile?

$X = \{ 2, 2, 3, 4, 5, 5, 5, 6, 7, 8, 8, 8, 8, 8, 99, 10, 11, 11, 12 \}$

$n = 20$.

$$\text{Index value} = \frac{99}{100}(n+1)$$

$$= \frac{99}{100}(21)$$

$$= 20.79$$

$\therefore n = 20$, the 99 percentile $= 12$

In Skewed data :-

Right skewed data
or
positively
Skewed data

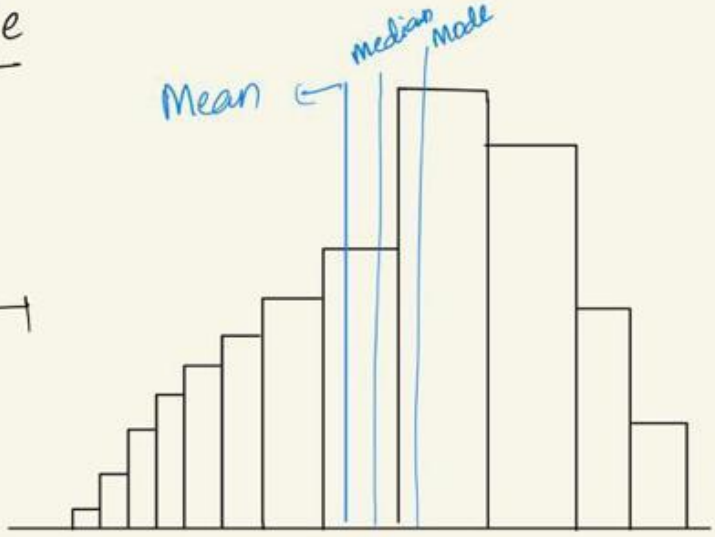
→ mean > median > mode

median is closer to 1st Quartile.

Left skewed data / negatively skewed data :-

→ mean < median < mode

median is closer to 3rd Quartile



5) Relationship between Mean, Median and mode in left and Right skewed data?

→ Illustrated the relationship using data < left skewed / Right skewed

Considering a data set ⇒ {1, 4, 5, 6, 9, 10, 13}.

$$\text{mean} = \mu = \frac{\sum x_i}{N} = 6.857$$

median = 6

} mean > median

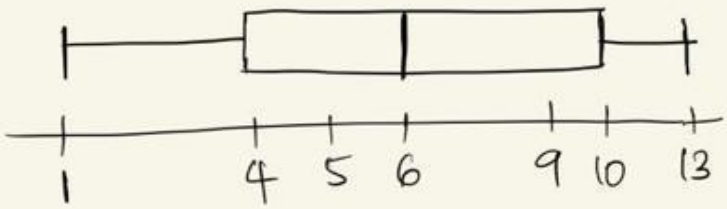
Box plot of given data :-

min value = 1

max value = 13

$$Q_1 = \frac{25}{100}(7+1) = \frac{1}{4} \times 8 = 2 \text{ index}$$

$Q_1 = 4$



$$Q_3 = \frac{75}{100}(8) = \frac{3}{4} \times 8 = 6 \text{ index}$$

$Q_3 = 10$.

Here median is closer to 1st Quartile

⇓

Data skewed right

Mode = 3 Median − 2 Mean

= 4.286.

⇒ mean > median > mode

Lower fence = $Q_1 - 1.5(IQR)$

= $4 - 1.5(10-4)$

= -5

Higher fence = $Q_3 + 1.5(IQR)$

= $10 + 1.5(6)$

= 19.

⇒ No outlier

Considering another data set ⇒ {8, 34, 35, 42, 43, 51, 60}

$\mu = 39$

median = 42

mode = 3(42) − 2(39)

mean < median < mode

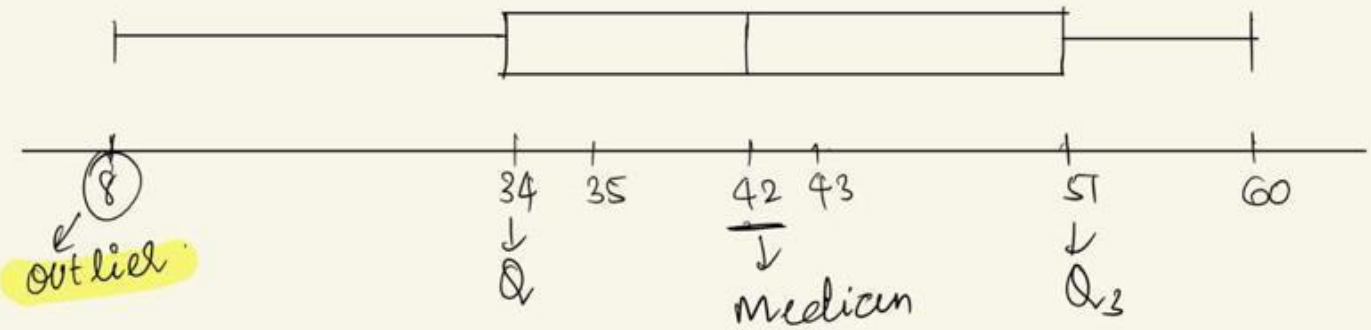
IQR = 51 − 34 = 48.

L.F = 8.5

H.F = 76.5

$Q_1 = \frac{25}{100} \times (8) = 2$ index $Q_1 = 34$

$Q_3 = \frac{75}{100} (8) = 6$ index $Q_3 = 51$



When there is an outlier ⇒ if left whisker is longer then the data is -vely skewed or left skewed data.