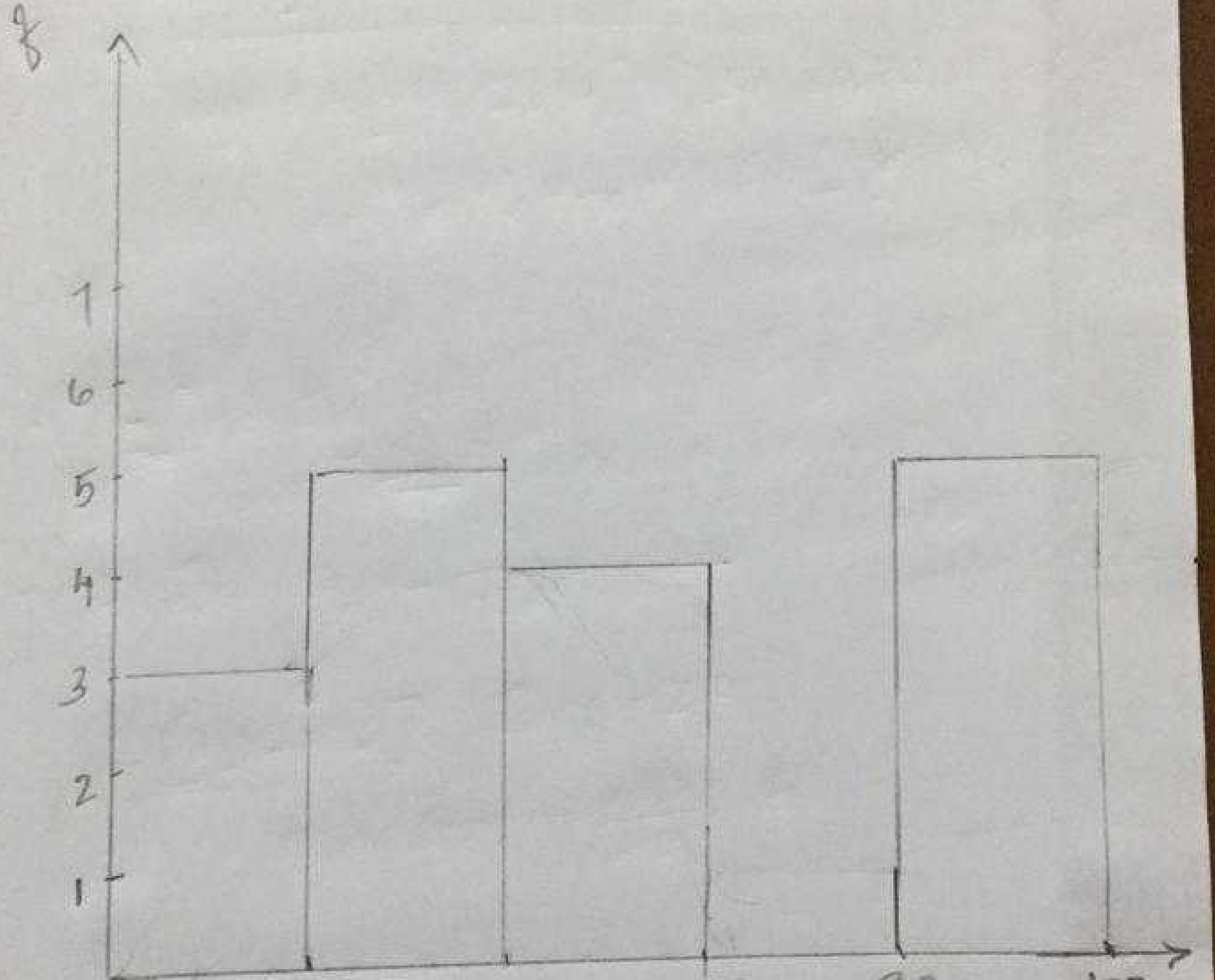
① Draw a histogram using the below data
{10, 13, 18, 22, 27, 32, 38, 40, 45, 51, 56, 57, 88, 90, 92, 94, 99}   bins = 5
binsize = 20



II In the quant test of CAT exam, the population standard deviation is known to be 100. A sample of 25 test takers has a mean of 520. Construct a 80% C.I about population mean.

If population standard deviation given, use Z test

$\sigma = 100 \quad n = 25$

$\bar{x} = 520$

σ = population S.D

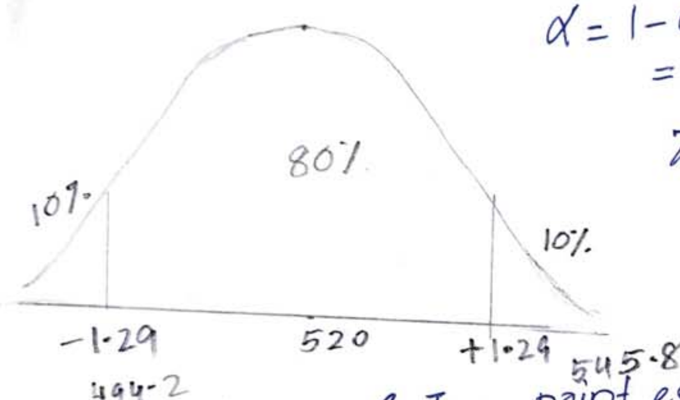
n = sample size

$\bar{x}$ = sample mean

$\alpha = 1 - 0.80$
$= 0.2$

α = Significance value

confidence Interval = 80%



$Z_{\alpha/2} = Z_{0.2/2}$
$= Z_{0.1}$

$Z_{0.1}$ = using Z table
$= 1 - 0.1 = 0.90 = 1.29$

C.I = point estimate ± margin error

$$= \bar{x} \pm Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$$

$$= 520 \pm Z_{0.1} \times \frac{100}{\sqrt{25}}$$

$$= 520 \pm 1.29 \times \frac{100}{\sqrt{25}}$$

lower fence $= \bar{x} - Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$

$$= 520 - 1.29 \times \frac{100}{\sqrt{25}}$$

$$= 494.2$$

Higher fence $= \bar{x} + Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$

$$= 520 + 1.29 \times \frac{100}{\sqrt{25}}$$

$$= 545.8$$

C.I = (494.2, 545.8)

IV A car company believes the percentage of residents in city ABC that owns a vehicle is 60% or less. A sales manager disagrees with this and he conducts a hypothesis testing surveying 250 residents and found that 170 responds yes to owning a vechicle.

a. State null hypothesis and alternate hypothesis

b. At 10% significance level, is there enough evidence to support the idea that vechicle ownership in city ABC is 60% or less ?

① Null hypothesis, $H_0: P_0 \leq 60\%$

Alternate hypothesis, $H_1: P_0 > 60\%$

$n = 250$

$x = 170$

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

$= 0.68$

$q_0 = 1 - P_0 = 1 - 0.6$

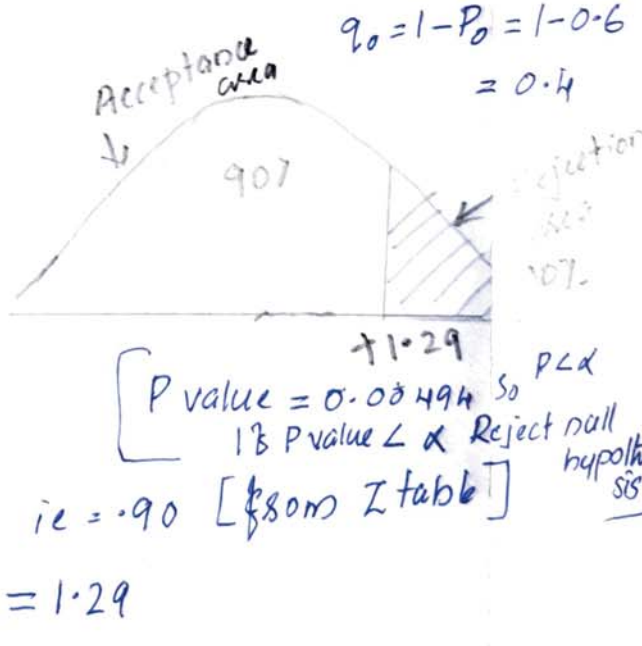
$= 0.4$

② Singnificance value, $\alpha = 10\%$ ie,

$\alpha = 0.1$

$C.I = .90$

③ Decision boundary



It is a one tail test because need to find one condition

Here $\alpha = 0.10$

$C.I = 90\%$ ie $= .90$ [from Z table]

$= 1.29$

[P value = 0.00494 So $P < \alpha$
If P value $< \alpha$ Reject null hypothesis]

④ Z test statistics

$$Z \text{ test} = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0 q_0}{n}}} = \frac{0.68 - 0.60}{\sqrt{\frac{.60 \times 0.40}{250}}} = \frac{0.08}{0.0309} = 2.5889$$

$2.5889 > 1.29$

⑤ Conclusion

$Z = 2.5889$. So 10% significance value is enough evidence to discard the null hypothesis.

Decision rule reject null hypothesis because $Z = 2.5889$ is $>$ than 1.29

4. What is the value of the 99 percentile?

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

$$\text{Percentile} = \frac{\text{Percentile} \times (n+1)}{100} \qquad n = 20$$

$$99 \text{ percentile} = \frac{99 \times 20+1}{100}$$

$$= \frac{99 \times 21}{100} = \underline{\underline{20.79}}$$

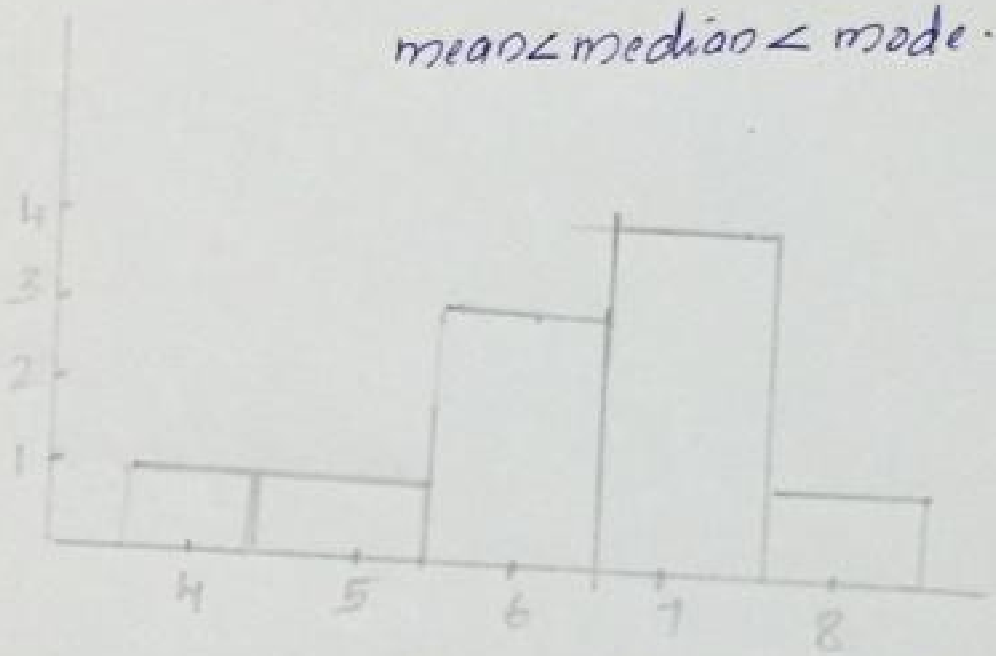
Assigment:

What is the relationship between mean, median and mode in left and right skewed distribution?

The mean is less than the median and they are both less than the mode in the distribution skewed to the left.

eg: 4 5 6 6 6 7 7 7 7 8 histogram for this data is.

mean < median < mode.



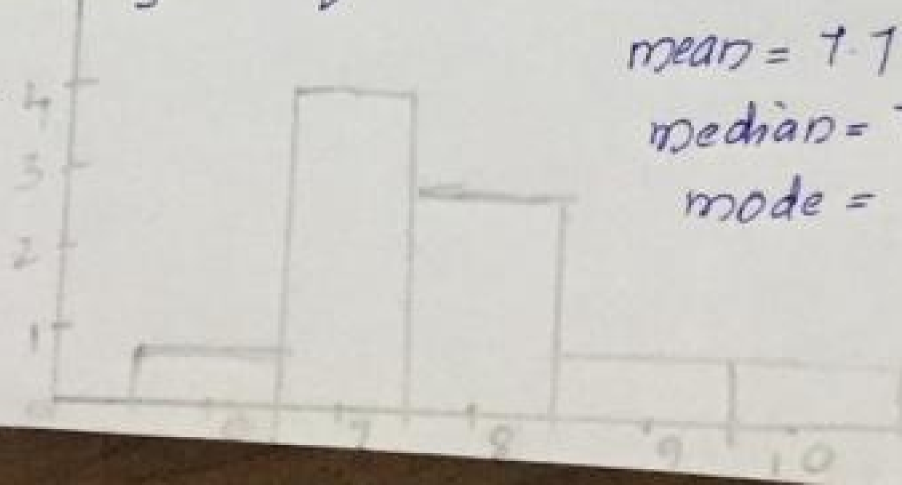
The mean is 6.3, the median is 6.5 and mode is 7.

The mean is the largest, while the mode is the smallest in the distribution skewed to the right.

eg: The histogram for the data 6 7 7 7 7 8 8 8 9 10

mean > median > mode

mean = 7.7
median = 7.5
mode = 7



III A company has 100K employees. Manager decided to buy Tshirt for them. He took a sample 500 and done a survey to know the numbers of Tshirt need to buy. Around 300 need Large size Tshirt and 200 need XL size Tshirt. Calculate the no. of Tshirts need to buy for 100K ?

Lets assume $C.I = 95\%$

$\alpha = 0.05$

$Z_{\alpha/2} = Z_{0.025}$

$= 1.96$ (from Z table)

$x_1 = 300$ $\quad x_2 = 200$

$n = 500$

Confidence Interval for a proportion is given by

$$C.I = \text{Sample proportion} \pm Z_{\alpha/2} \times \sqrt{\frac{\text{sample proportion} \times (1 - \text{sample proportion})}{n}}$$

Assume P = Sample proportion, So

$$C.I = P \pm Z_{\alpha/2} \times \sqrt{\frac{P(1-P)}{n}}$$

Cacluate C.I for Large size Tshirt,

$$L.F = P - Z_{\alpha/2} \times \sqrt{\frac{P(1-P)}{n}}$$

$$P = x_1/n = \frac{300}{500} = 0.6$$

$$L.f = .6 - 1.96 \times \sqrt{\frac{.6 \times (1-.6)}{500}}$$

$$= .6 - 1.96 \times \sqrt{\frac{.6 \times .4}{500}}$$

$$= .6 - .0009408 = \cancel{.5990}\ .5510$$

$$H.F = P + Z_{\alpha/2} \times \sqrt{\frac{P(1-P)}{n}} = .6 + 1.96 \times \sqrt{\frac{.6 \times .4}{500}}$$

$$\cancel{= 0.6009} = 0.6429$$

C.I for large Tshirt is $(.5990, 0.6009)$

$= (.5510, 0.6429)$

Calculate C.I for xlarge T.shirt.

$C.I = 95\%$

$x = 200$ $\quad n = 500$

$P = x/n = \frac{200}{500} = .4$ $\quad Z_{\alpha/2} = 1.96$ from table

$$L.F = P - Z_{\alpha/2} \times \sqrt{\frac{P(1-P)}{n}}$$

$$= .4 - 1.96 \times \sqrt{\frac{.4(1-.4)}{500}}$$

$$= 0.4 - 1.96\sqrt{\frac{0.4 \times 0.6}{500}} = \cancel{0.34905}$$

$$= 0.3570$$

$$H.F = P + Z_{\alpha/2} \times \sqrt{\frac{P(1-P)}{n}} = .4 + 1.96\sqrt{\frac{.4 \times .6}{500}}$$

$$= 0.4429$$

For large Tshirt the C.I=95% with proportion of sample 500 is [.5570, .6420] ie approximately 55000 to 64000 ~~T~~ Tshirts need to buy for Large size for 100k

For extra large Tshirt with C.I=95% with proportion of sample 500 is [.3570, 0.4429] ie approximately 35000 to 44 thousand XL T.shirts need to buy for 100k employees

I What is the average size of all sharks in the world?

This problem can be solved using inferential statistics. Lets assume the following values,

σ = standard deviation, n = sample

$\sigma = 100$ $n = 50$

sample mean, $\bar{x} = 300$ Confidence Interval = 95%

α = significance value

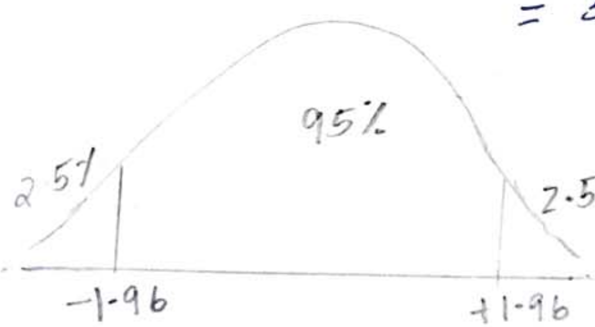
$\alpha = 0.05 = 1 - .95 = 0.05$

$C.I = \text{Point estimate} \pm \text{margin error}$

$= \bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

$= 300 \pm Z_{0.05/2} \times \frac{100}{\sqrt{50}}$

$= 300 \pm 1.96 \times \frac{100}{\sqrt{50}}$ $Z_{\frac{0.05}{2}} = Z_{0.025}$



$Z_{0.025}$ find the values using Z table

~~Z=0.~~ $Z_{0.025} = 1 - 0.025 = 0.975$

$= 1.96$

Calculate lower limit $= 300 - 1.96 \times \frac{100}{\sqrt{50}}$

$= 272.277 = 272$

Higher limit $= 300 + 1.96 \times \frac{100}{\sqrt{50}}$

$= 327.7 = 328$

With the 95% confidence Interval population mean is in between 272 to 328