

ASSIGNMENT - 1

Statistics

1. Plot a histogram

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

Ans:

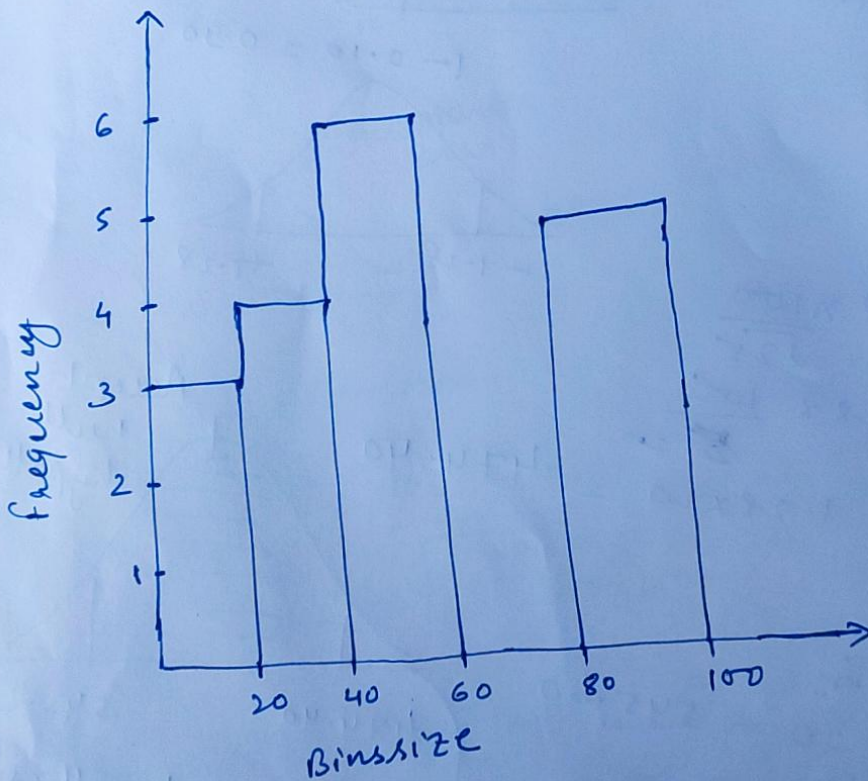
Data is already arranged in ascending order

Let's say bins given = 5

Range as we can see = 0 to 100

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

Plotting histogram:



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 car.

a) State the null hypothesis & alternate.

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

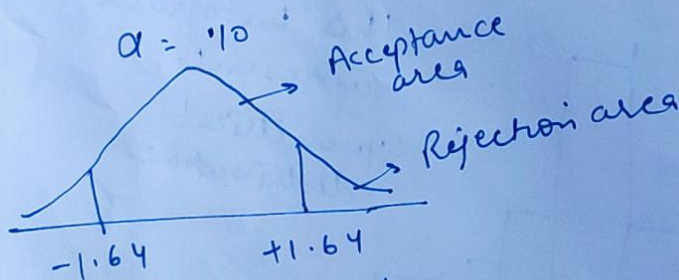
$$n = 250$$

$$n = 170$$

Ans. $H_0: P_0 = 60\%$
 $H_1: P_0 < 60\%$

$$\hat{P} = \frac{n}{n} = \frac{170}{250} = \frac{17}{25} = 0.68$$

$$P_0 = ? \quad q_0 = 1 - 0.60 = 0.4$$



Z-Test with proportion

$$\frac{\hat{P} - P_0}{\sqrt{\frac{P_0 q_0}{n}}} = \frac{0.68 - 0.60}{\sqrt{\frac{0.60 \times 0.4}{250 \times 1000}}} = \frac{0.08 \times 1000}{1.009 \times 100} \Rightarrow \frac{80}{9}$$

= 8.8
 So we reject the null hypothesis
 citizens own a vehicle is less than 60%.

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

Ans
 \Rightarrow Value = $\frac{99}{100} \times (20+1)$

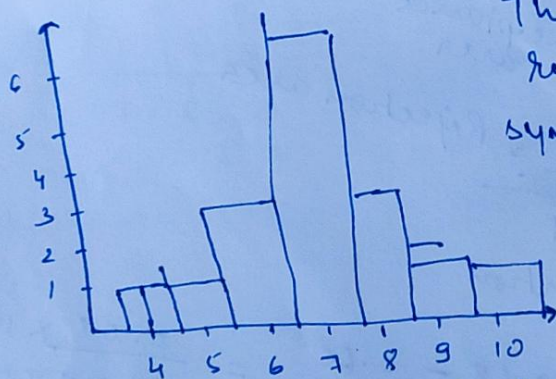
$$\frac{99}{100} \times 21 \Rightarrow 20.79$$

\therefore 20 is our index

12 is 99 percentile value

5) In left & right skewed data, what is the relationship b/w mean, median & mode?
Draw the graph to represent the same.

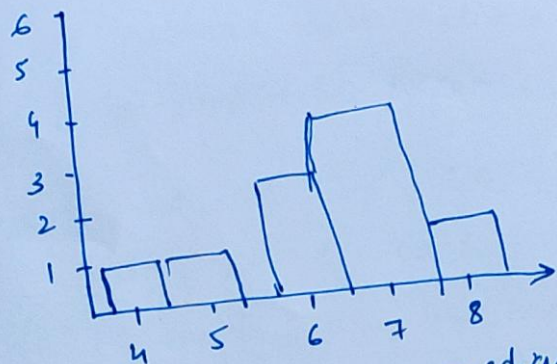
Ans: Consider a data set
4, 5, 6, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 8, 9, 10
Plotting histogram



This histogram represents symmetrical distribution

So in perfectly symmetrical distribution, the mean & the median are same

Considering another distribution
(4, 5, 6, 6, 6, 7, 7, 7, 7, 8)

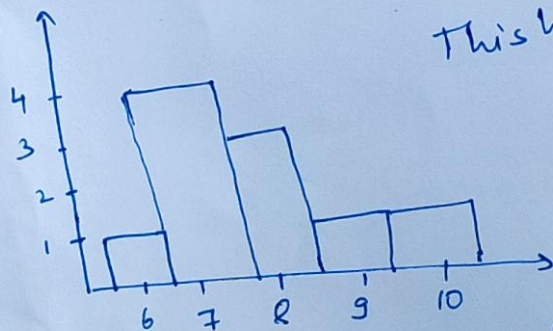


Here if we
calculate $\text{mean} = 6.3$
• $\text{median} = 6.5$
 $\text{mode} = 7$

$\text{mean} < \text{median}$
 $\text{mean} < \text{median} < \text{mode}$
This histogram is left skewed.

Consider the data set
6, 7, 7, 7, 7, 8, 8, 8, 9, 10

Plotting
histogram



This histogram is right
skewed.

$\text{Mean} = 7.7$
 $\text{Median} = 7.5$
 $\text{Mode} = 7$

mean is largest
 mode is smallest

Other Assignment Questions

1) log normal distribution problem.
Let us consider incomes of people

20, 22, 25, 30, 60

Taking log of each income

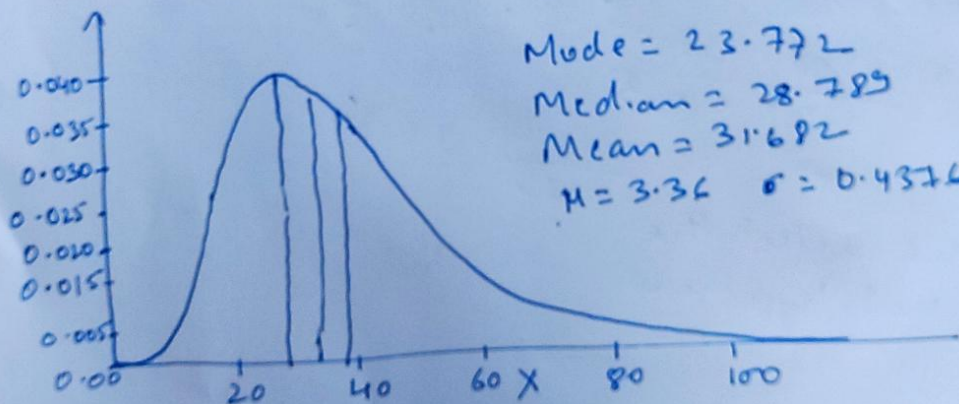
income	$\ln(\text{income})$	μ	$\ln(\text{income} - \mu)$	$\ln(\text{income} - \mu)^2$
20	2.9957	3.36	-0.3643	0.1327
22	3.0910	3.36	-0.269	0.0723
25	3.2188	3.36	0.4412	0.0193
30	3.4011	3.36	0.0411	0.0017
60	4.0943	3.36	0.7343	0.5393

$$\hat{\mu} = \frac{2.99 + 3.09 + 3.21 + 3.40 + 4.09}{5} = 3.36$$

$$\sigma = \sqrt{\frac{\sum (\ln x - \hat{\mu})^2}{n-1}}$$

$$\sigma = \sqrt{\frac{0.1327 + 0.0724 + 0.0193 + 0.0017 + 0.5393}{5-1}}$$

$$\Rightarrow 0.4376$$



2) T-shirt problem

100k employees
500 samples taken

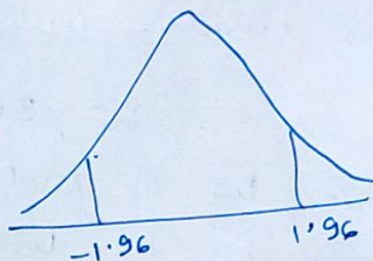
CE = 95%

$$H_0: L = XL$$

$$H_1: L \neq XL$$

$$\alpha = 0.05$$

$$L \& XL = 300 \& 200$$



Test statistics

$$\hat{p}_1 = \frac{300}{500} = 0.60$$

$$\hat{p}_2 = \frac{200}{500} = 0.40$$

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$\hat{p} = \frac{200+300}{500+500}$$

$$\Rightarrow \frac{500}{1000}$$

$$\Rightarrow 0.50$$

$$\frac{0.60 - 0.40}{\sqrt{0.50(1-0.50) \left(\frac{1}{500} + \frac{1}{500} \right)}}$$

$$\Rightarrow \frac{0.20}{\sqrt{0.50(0.50) \left(\frac{2}{500} \right)}}$$

$$\frac{0.20 \times 1000}{0.50 \times 0.063}$$

$$\Rightarrow \frac{2000}{5 \times 63} \Rightarrow 6.34$$

\therefore we reject the null hypothesis