JHALAK

DATA ANALYTICS (PSDA)

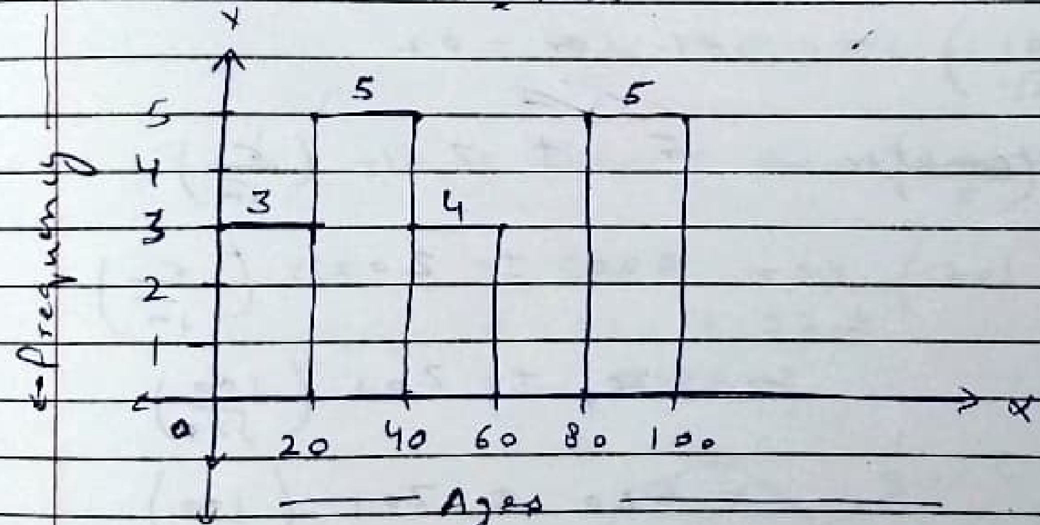
ASSIGNMENT-1

Histogram

Q.1 Eg. Ages = { 10, 13, 18, 22, 27, 32, 38, 40, 45, 51, 56, 57, 88, 90, 92, 94, 99 }

Bins = 5
Bin size = 20

Sol. C.I. = (0-20), (20-40), (40-60), (60-80), (80-100)
Class Interval



Q.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 a 80% Confidence Interval about the mean.

Sol. Population standard deviation (σ) = 100
Sample size (n) = 25
Sample mean ($\bar{x}$) = 520
(C.I) Confidence Interval = 80%

As, population standard deviation is given we will apply z-test.

Parameter for confidence Interval = Point Estimate ± Margin of Error.

$$\text{Parameter} = \bar{x} \pm Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

α = Significance Value
= 1 − C.I.
= 1 − 80%
= 20% or 0.2

$$\text{Parameter} = \bar{x} \pm Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$= 520 \pm Z_{0.2/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$= 520 \pm Z_{0.1}\left(\frac{100}{\sqrt{25}}\right)$$

$$= 520 \pm Z_{0.1}\left(\frac{100}{5}\right)$$

$$= 520 \pm Z_{0.1}\,(20)$$

We'll match the value of (1 − 0.1) i.e. 0.9 from z-table = $\boxed{1.39}$

For Confidence Interval,
We need to find Lower fence & Higher fence

① Lower fence = Point Estimate − Margin of Error

$= \bar{x} - Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

$= 520 - Z_{0.1}\left(\frac{100}{\sqrt{25}}\right)$

$= 520 - 1.39(20)$

$= 520 - 27.8$

$= 492.2$

② Higher fence $= \bar{x} + Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

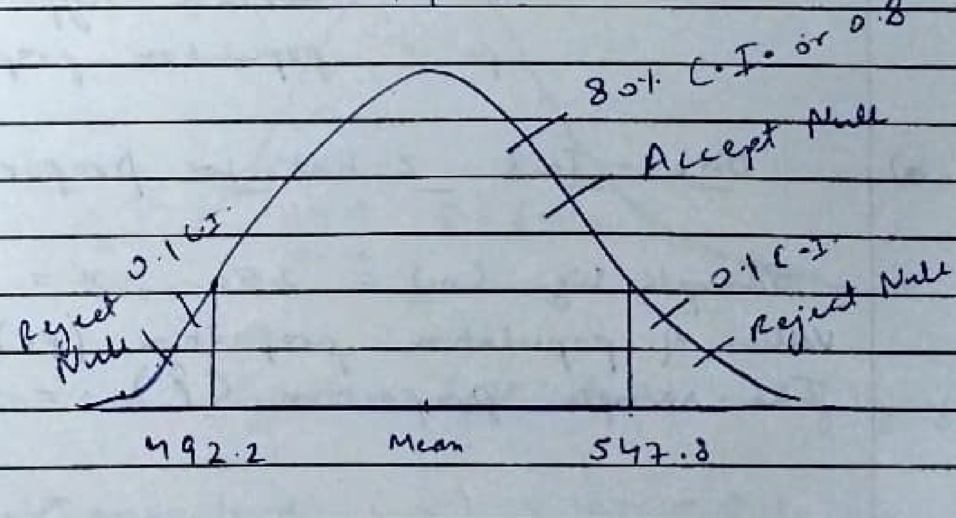
$= 520 + Z_{0.2/2}\left(\frac{100}{\sqrt{25}}\right)$

$= 520 + Z_{0.1}\left(\frac{100}{5}\right)$

$= 520 + 1.39(20)$

$= 520 + 27.8$

$= 547.8$



Normal Distribution
or
Probability Density function

Q-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 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.

Sol. a) H_0: Percentage of citizens in city ABC that owns a vehicle (p_0) $\leq$ 60%.

H_1: Percentage of citizens in city ABC that owns a vehicle (p_0) $>$ 60%

where, H_0 = Null Hypothesis
H_1 = Alternate Hypothesis
p_0 = population proportion

b) One - Tail Z-test for proportions:

Sample size (n) = 250, x = 170

Value of population proportion (p_0) = 60%

The sample proportion ($\hat{p}$) $= \frac{x}{n} = \frac{170}{250}$

$= 0.68$

Calculation of Z-Statistics:

$$Z = \frac{\hat{p} - p_0}{S.E.}$$

S.E. = Standard Error = $\sqrt{\frac{p_0(1-p_0)}{n}}$

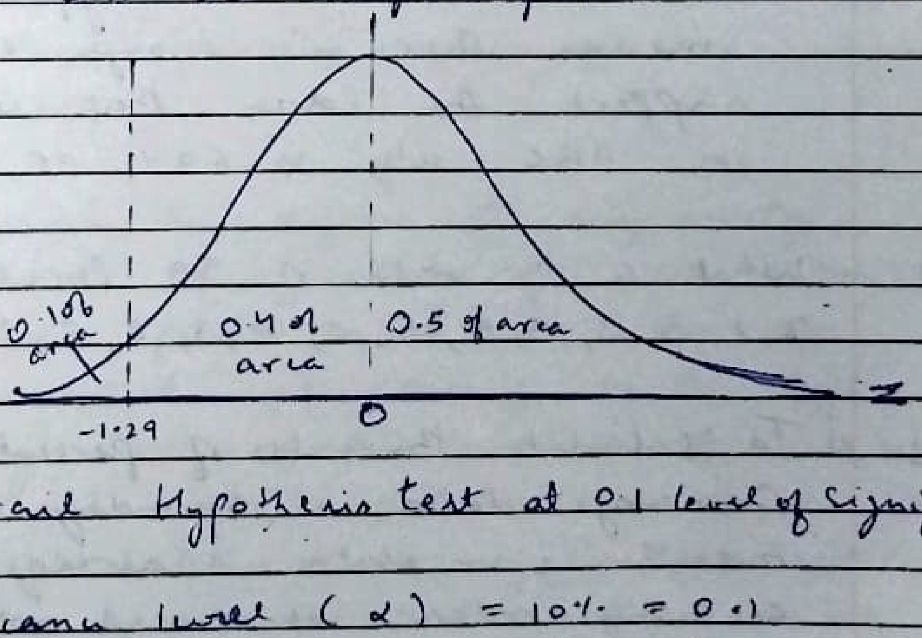
So, $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

$$z = \frac{0.68 - 0.6}{\sqrt{\frac{0.6(1-0.6)}{250}}}$$

$$z = \frac{0.08}{\sqrt{\frac{0.6(0.4)}{250}}}$$

$$z = \frac{0.08}{0.03}$$

Z or Standardised Sample Proportion = 2.667



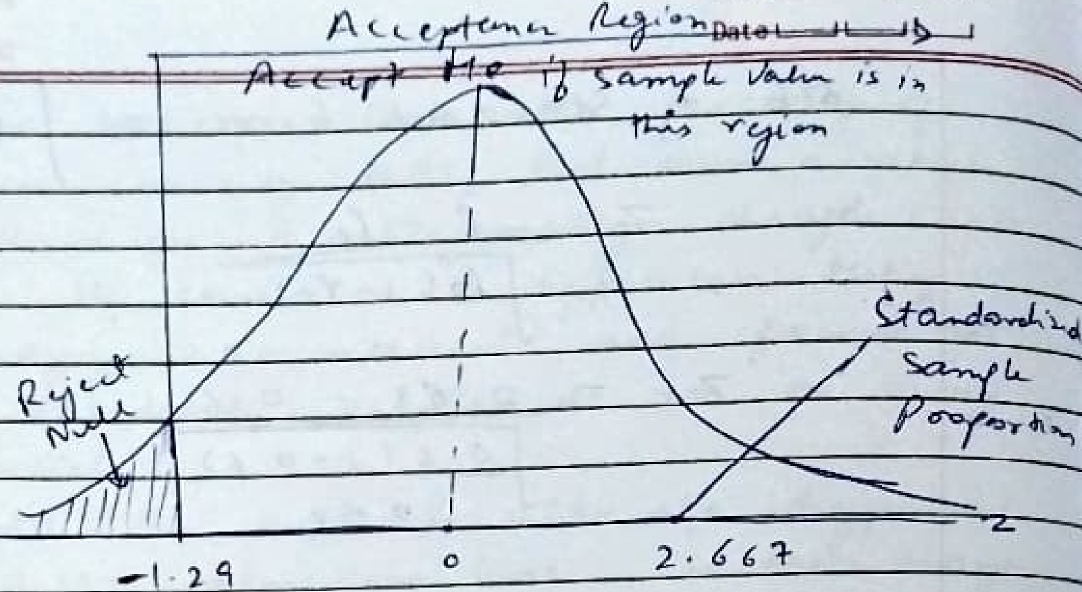
One-tail Hypothesis test at 0.1 level of Significance

Significance level (α) = 10% = 0.1

$Z_{0.1}$ = Value of Z at (1-0.1)

= Value of Z at 0.9 → Match with Z-table

= -1.29



One-tailed (left-tailed) Hypothesis test at 0.1 level of significance showing the acceptance region and standardised sample proportion.

P-Value $= Z_{2.667} = 0.99621$

A. ~~S~~tandardised sample proportion $> Z_{0.1}$

i.e. $2.667 > -1.29$

we'll accept null hypothesis which means there is enough evidence to support the idea that vehicle owner in ABC city is 60% or less.

Q-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

Sol ① To calculate the value of percentile, first arrange data in ascending order but the given data is already arranged in ascending order, we ~~d~~ need not do it.

② Value of 99 percentile

$$= \left(\frac{\text{Percentile}}{100}\right)(n+1)$$

where n = number of values in dataset

$$99 \text{ percentile} = \left(\frac{99}{100} \times (20+1)\right)$$

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

$$= (20.79)^{th} \text{ Value}$$

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

Value of 99 Percentile $= \boxed{11.5}$

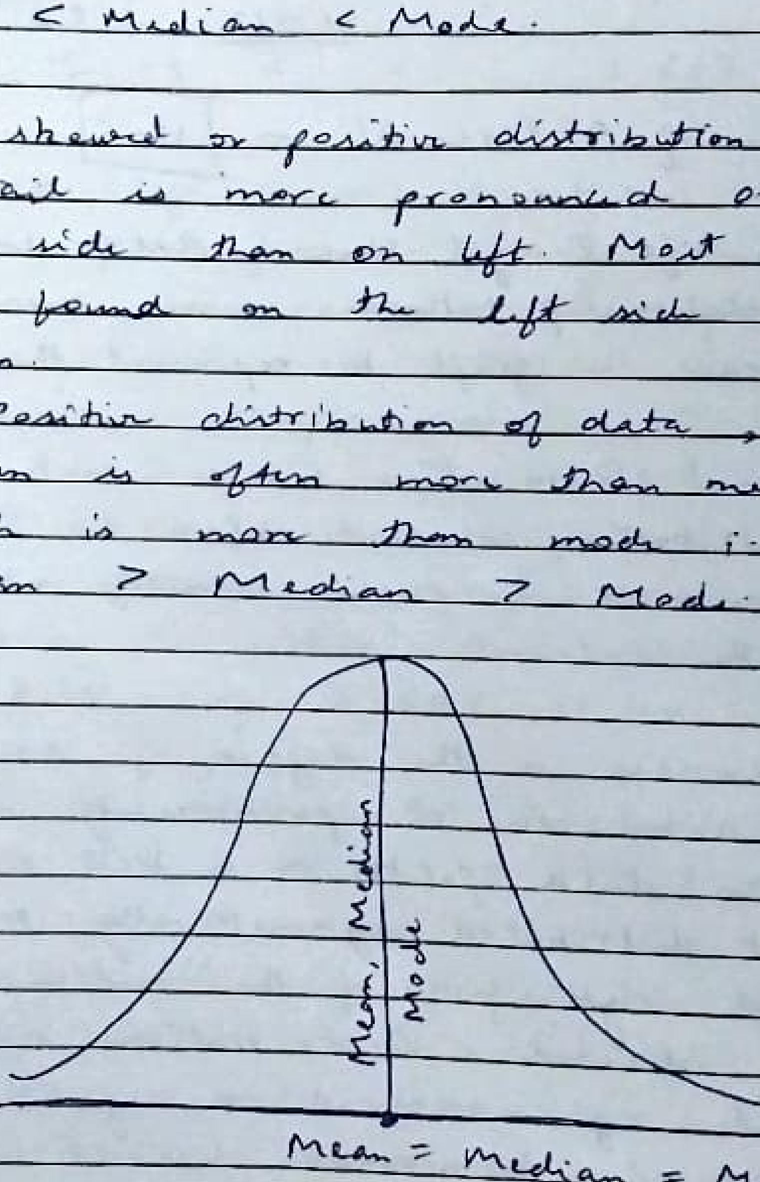
Q-5 In left & right skewed data, what is the relationship between mean, median & mode? Draw the graph to represent the same.

Sol: Normal Distribution is a continuous probability distribution wherein values lie in a symmetrical fashion mostly situated around the mean.

Skewness is the degree of asymmetry observed in a probability distribution. When data points on a bell curve are not distributed symmetrically to the left and right sides of the median, the bell curve is skewed. Distributions can be positive and right-skewed or negative and left-skewed. A normal distribution exhibits zero skewness.

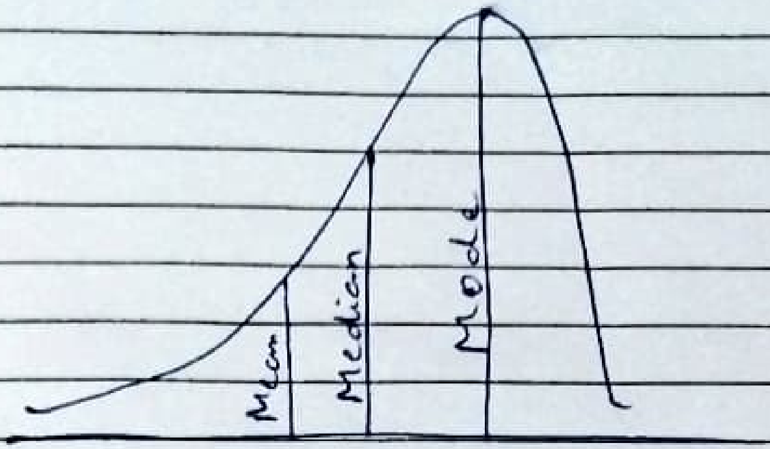
Negative or left-skewed reffers to a longer or fatter tail on the left side of the distribution. Most values are found on the right side of the mean.
In negative distribution of data, the mean is less than the median, which is often less than the mode i.e. Mean < Median < Mode.

Right-skewed or positive distribution means its tail is more pronounced on the right side than on left. Most values are found on the left side of the mean.
In positive distribution of data, the mean is often more than median, which is more than mode i.e. Mean > Median > Mode.



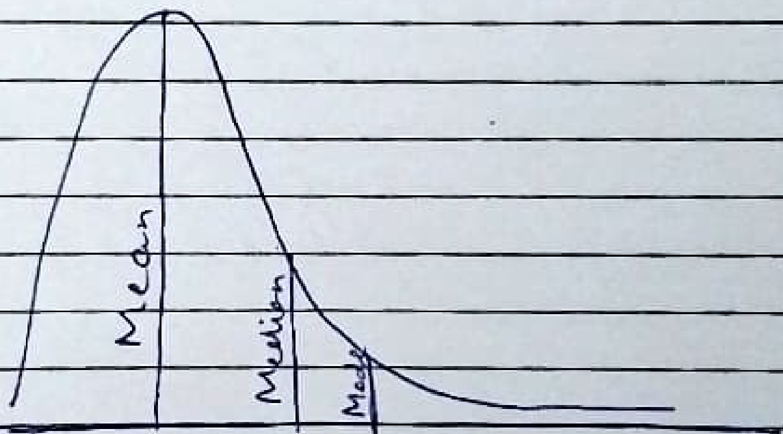
Mean = Median = Mode

Normal Distribution



Mean < Median < Mode

Negative Distribution



Mean > Median > Mode

Positive Distribution