Descriptive Statistics

8/12/23

Propability → Why?

We analyze, understand and shape of the data we see A to Z of the data in Statistics.

But when it comes to probability we predict or forecast the data.

Why Probability & Statistics → Real World

It is used for decision Making.

Descriptive Statistics

Statistical functions that describe the data

The first step in this is EDA

Exploratory data Analysis

→ Mean → Sum of elem / No of items

→ Median – Middle Most element

~~Even~~ odd

$\frac{n+1}{2}$

~~odd~~ Even

$\frac{\frac{n}{2} + \left(\frac{n}{2}+1\right)}{2}$

→ Mode - frequently Occured number

Mean – Average

Median – Tells ~~the~~ about how 50% of data arranged

for Example Here Sales

10 10 10 20 40 60 70 80 80

Here Median is 40

So here 50% of people made the Sales with in 40 Rs. and other half made a sale which is greater than ₹40

Mode :– How many times it occurs in data

Quiz:–

(1) Median – 10 20 30 40 50 60 70

(2) Median – 10 20 30 40 50 60 70 80

$\frac{n}{2}+\left(\frac{n}{2}+1\right)$ → 45

$\frac{n+1}{2}$ (40, 50)

$$= \frac{\frac{8}{2}+\left(\frac{8}{2}+1\right)}{2}$$

$$= \frac{4+5}{2} = \frac{40+50}{2} = 45$$

(3) 4 people → Average is 24

Three people ~~Average~~ age – 20, 22, 26

Median age ?

$$\frac{20 + 22 + 28 + x}{4} = 24$$

$$70 + x = 96$$

$$x = 96 - 70$$

$$x = 26$$

20, 22, 26, 28

$\left(\frac{n}{2}, \frac{n}{2}+1\right)$ $(2, 3)$ $\left(\frac{22+26}{2}\right)$

Median age **24**

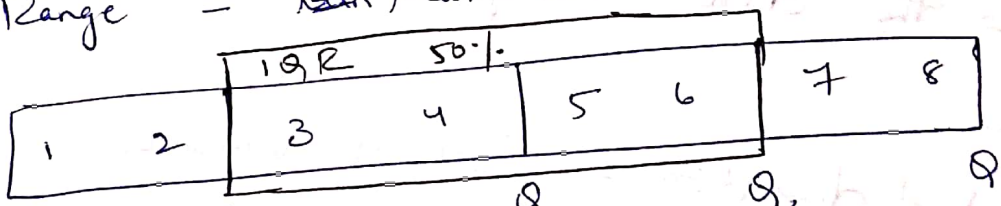
④ Mean weight of 2 child = 40 kgs

If weight of mother included = 45

Weight of mother ?

$$\frac{40 + x}{2} = 45 \qquad x = 50$$

→ Range = Max − Min



Q_1 25% Quartile 1

Q_2 50% Quartile 2

Q_3 75% Quartile 3

Q_4 100% Quartile 4

Quartiles helps us understand how the data is distribution.

One of the important Measure is

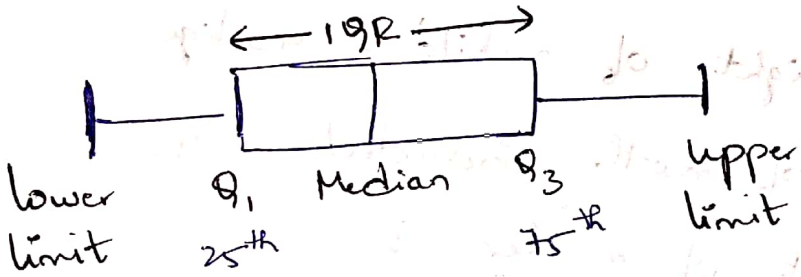
IQR Inter Quartile Range

$$IQR = Q_3 - Q_1$$

How 50% of Middle data distributed

Upper Limit in IQR = $Q_3 + 1.5 * IQR$

Lower Limit in IQR = $Q_3 - 1.5 * IQR$



- Variance → Measure average Squared deviation of each data point from the mean

How the data is spread according to the mean

- Standard deviation — Root of Variance

Kurtosis

The measure of Peakedness of data

+ve kurtosis

Leptokurtic

thin

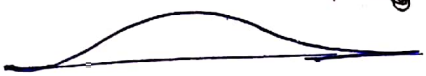
Normal distribution

Mesokurtic

-ve kurtosis

flat

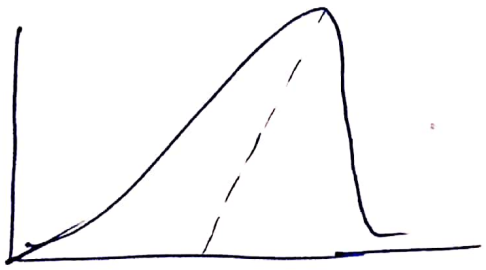
Platykurtic



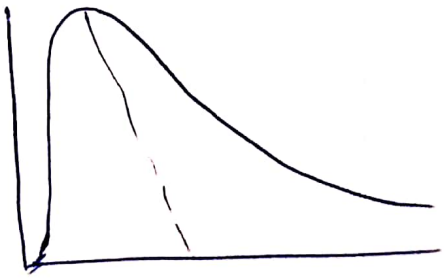
Charatesitic

Skewness

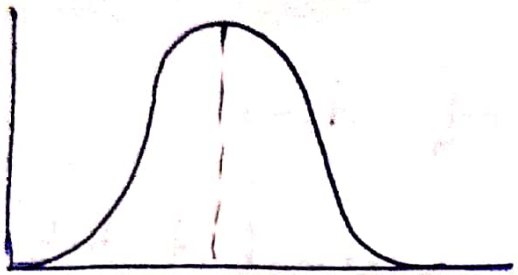
Understanding the spread of data with respect to kurtosis. [Skewness of data]



Negative Skew



Positive Skew



Bell Shaped Curve

Not Skewed

Skewness measure the asymmetry of a distribution of the data