Measure of Dispersion / Variability

3 Methods to calculate :-

1. Variance
2. Standard deviation
3. Inter-Quartile range
4. Coefficient of variation
5. Range = $X_{max} - X_{min}$

Variance $\frac{\Sigma(x - \bar{x})^2}{n}$

→ average of the squared differences of the observations from their mean

Sample variance :-

$$s^2 = \frac{\Sigma (X - \bar{x})}{n}$$

Population variance

$$\sigma^2 = \frac{\Sigma (X - \mu)^2}{N}$$

n : Sample size N : Population size

(ii) Another formula :- (Variance)

$$\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2$$

(iii) Another formula :-

$$\frac{\Sigma (\bar{X} - \bar{x})^2}{n-1}$$

Standard Deviation :-

① $s = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n}}$ $\sigma = \sqrt{\frac{\Sigma (x-\mu)^2}{N}}$

μ = population mean

②

$s = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$ $\sigma = \sqrt{\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2}$

③

Note :- Just take square root of variance to take Std. Deviation

IQ Range

$$Q_3 - Q_1$$

IQR = upper quartile - lower quartile

→ The middle 50% of data
→ It only tells range (max - min) of the middle range of data.

→ Can only calculate dispersion of the middle half of the data.

Coefficient of variation :—

$$(CoV)_N = \frac{\sigma}{\mu} \times 100 \quad \text{(population)}$$

$$(CoV)_n = \left(\frac{s}{\bar{x}} \times 100\right) \quad \text{(sample)}$$

→ It is called relative measure of coefficient

→ To compare two datasets whose value are of different unit

→ relative measure of dispersion

Q⇒ A home theatre in a box is easiest way to provide sound surround for a home entertainment center

A sample of prices is shown the prices are for models with a DVD player and for models without a DVD player.

DVD compatable	Price	Simple Model	Prices
	450		300
	300		300
	400		360
	500		290
	400		300

$\bar{x} = 310$

average of DVD = 410

std deviation = $\sqrt{\frac{\Sigma (x - \bar{x})^2}{N}}$

$= \sqrt{\frac{862500 - \frac{6050)^2}{25}}{5}}$

$= \sqrt{172500 - 168100}$

$= 66.33$

$= 74.16198$

Std dev of simple $= \sqrt{\frac{483700}{5} - (310)}$

$\Rightarrow 36.8781$

Wallpole & - 1.17 and 1.21

homework

Steps of SoD example

$$S_x = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

$$= \sqrt{\frac{\Sigma x^2}{n} - (\bar{X})^2}$$

$$= \sqrt{\frac{21921}{10} - (210)^2}$$

$$= 110.27$$

31 August 2023 EXAMPLE of CV

Life in hours and price in rupees : 2 datasets :-

Price PKR		Life (hrs)	
X	x^2	Y	y^2
8	64	130	16900
13	169	150	22500
18	324	180	32400
23	529	250	62500
30	900	345	119025

→ estimate dispersion in both datasets:-

$\bar{X}$ = 18.4 (PKR)

$\bar{X}_2$ = 211 (life hrs)

$$\text{S.D of } x = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma \bar{x}}{n}\right)^2}$$

$$= \sqrt{\frac{\Sigma x^2}{n} - (\bar{X})^2}$$

= 7.6576

S.D of Y = 78.383

CoV of x = 41.6 %

CoV of Y = 37.1 %

note:-
X has more disp. but Y has less disp. on basis of a
The 5% c.v is a good data.

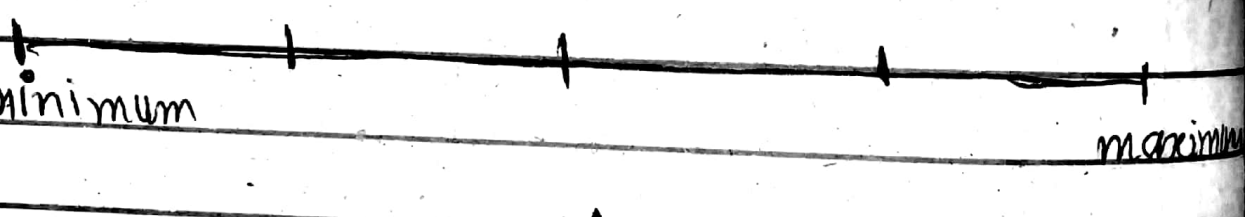
31 August 2023

IQ

Q3

FIVE POINT SUMMARY

1) Minimum
2) Median
3) Q_1
4) Q_2
5) Maximum



→ describe the data in 5 points.?

=upper

The

It o

of T

Can

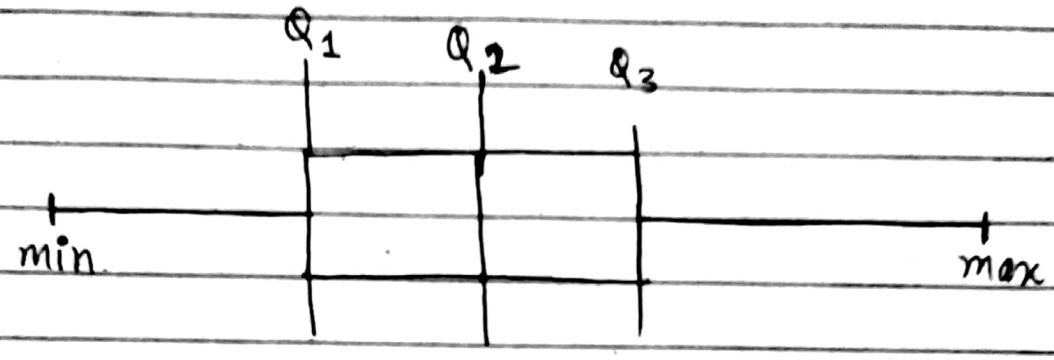
mid

oe.

(C.V)$_N$

C.V)$_n$

Box and Whiskers



→ It tell us about data distribution or the shape of data

→ It can tell us the skewness of data

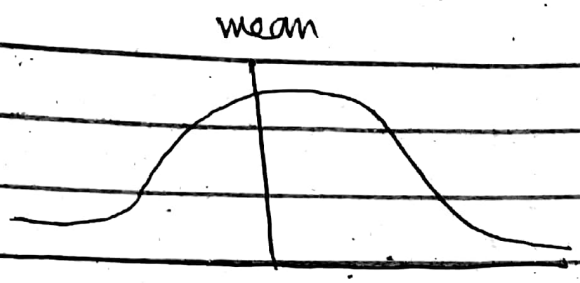
→ For example if right whisker is much greater than left whisker then data is positively skewed

→ If the left whisker is greater than right whisker the the data is negatively skewed

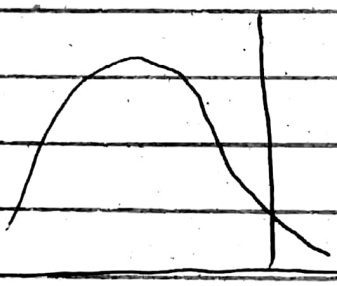
→ It can also tell outliers as well eg.

→ If length of right and left whisker then the data is symmetrical

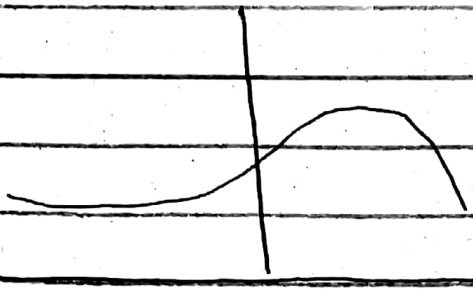
★ if data is symmetrical then mean, median and mode will be equal or same on a normal curve



Normal curve (bell-shaped)

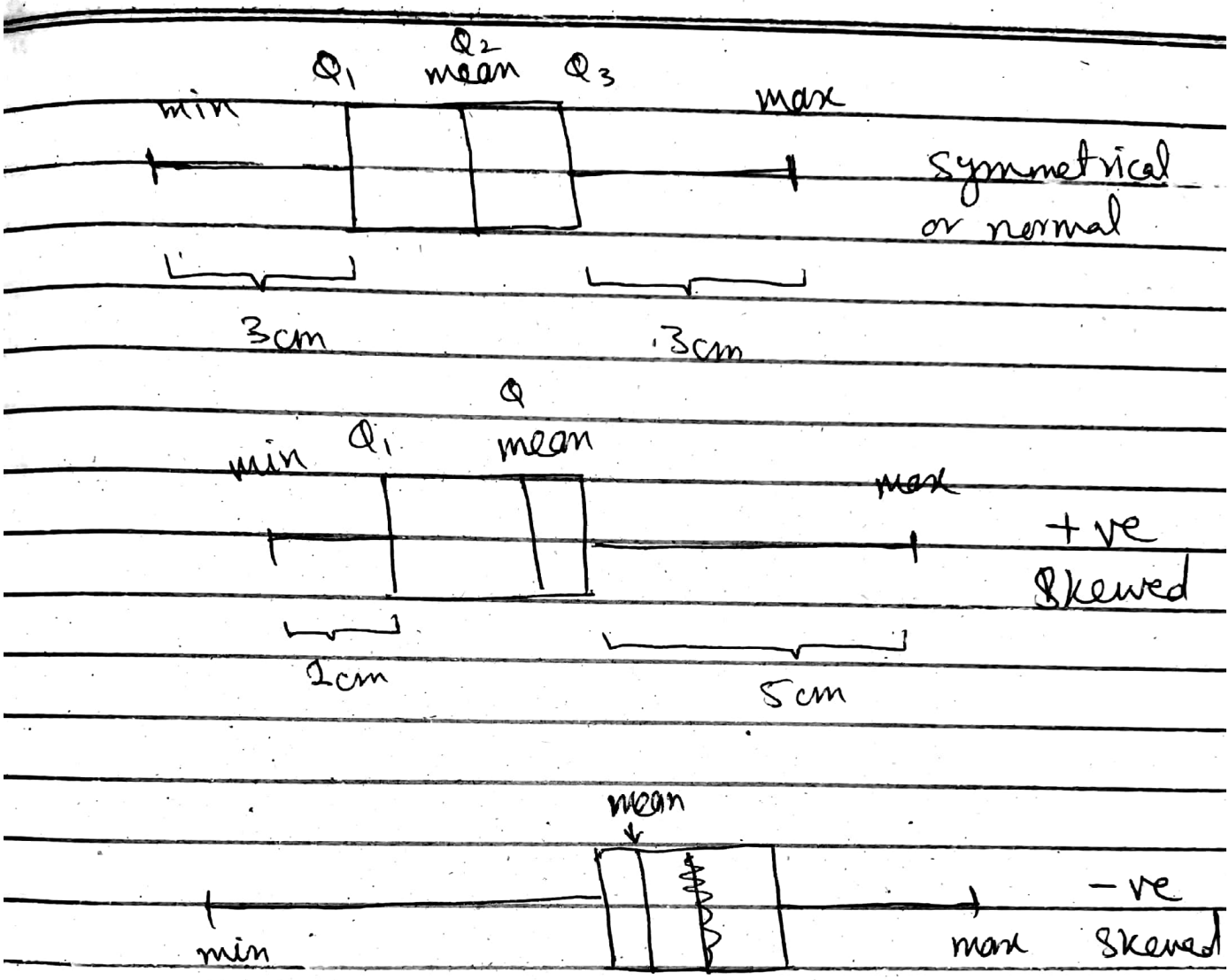


~~negatively~~ skewed (+ve)



~~positively~~ skewed (-ve)

★ graph analysis is subjective
★ so calculate fences, (limits)



mean = mode = median (symetry)

mean > median > mode (+ve skew)

mean < median < mode (-ve skew)

Calculate Fences

Lower limit = $Q_1 - 1.5\,(IQR)$

Upper limit = $Q_3 + 1.5\,(IQR)$

* Any data outside of upper or lower limit is is for sure an outlier -

* A data or a point may be called an outlier with this diagnostic test of Fences, but that will be difficult.

Example

→

Time in seconds for 12 vehicles to exit a parking lot in a specific location 8-

~~145~~	mean = 303.3
~~105~~	
260	
330	
250	
~~195~~	
375	

Sorted :

105	105
180	145
195	150
	180
	195
	250
	260
	380
	375
	420
	480
	750

$Q_1 = \frac{150 + 180}{2} = 165$

$Q_2 = 255 \quad (6^{th} + 7^{th})$

$Q_3 = 397.5 \quad (9^{th} + 10^{th})$

$= \frac{375 + 420}{2}$

mean = 303.3

median = 255

mode = x