05/01/21 Lec- 04

Topics to be discussed :

1) Mode (measure of central tendency)

2) dispersion (range, quantiles, quartiles, IQR, box-plots, variance, standard deviation, Z-score)

Steps to find mode -

1. find all the distinct categories / or values in your data set.
2. write down the frequency of each value or category.
3. Pick out the one('s) with the highest frequency to get the mode.

Eg: 1.

Value	1	2	3	4	5	6	7	8
f	4	6	4	4	3	2	1	1

$\boxed{\text{mode} = 02}$

2.

category	blue	red	green	pink	Yellow
f	4	5	8	1	3

$\boxed{\text{mode} = \text{green}}$

3.

Values	1	2	3	4	5
f	2	3	3	3	3

$\boxed{\text{mode} = 2, 3, 4, 5}$

* data sets with one / two / three modes are called unimodal, bimodal / trimodal resp.

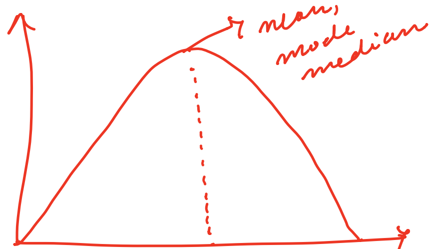
In general, data sets with 2 or more modes is called multi-modal.

Relation between mean, mode, median :

$$\boxed{(\text{mean} - \text{mode}) = 3 \times (\text{mean} - \text{median})}$$

NOTE :

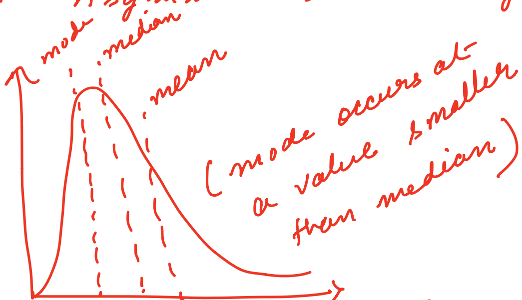
1. In a unimodal frequency, with perfect symmetric data distribn; the mean, mode and median are all at same center value.



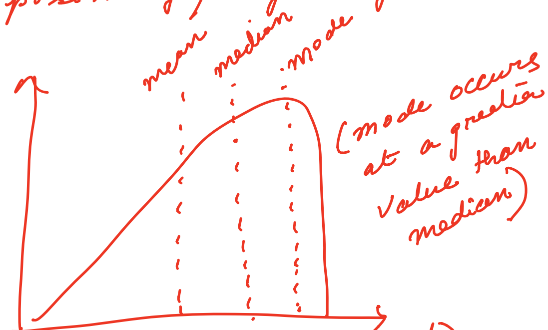
$\mu = 10$
mode $= 10$
median $= 10$

(symmetric data)

2. Asymmetric data may be positively / negatively skewed.



(positively skewed)



(negatively skewed)

→ Data dispersion :

defn : The degree to which the numerical data spreads is called dispersion / or variance of data.

① Range :

points score per game

player A :

player A	7	8	9	10	11	12	13
f	1	1	2	2	2	1	1

$\mu = 10$
Mode $= 10$
median $= 10$

player B

player B	7	9	10	11	13
	1	2	4	2	1

$\mu = 10$ mode $= 10$
median 10

player c :

player c	3	6	7	10	11	13	~~30~~

$\mu = 10$
mode $= 10$
[illegible] $= 10$

2 | 1 | 2 | 3 | 1 | 1 | 1 mean...

Range$_A$ = UB - LB
= 13 - 7
= 06

Range$_B$ = 13 - 7
= 06

Range$_c$ = 30-3 = 27
or 13 - 3
= 10

2) Quantiles → Quartiles → deciles → percentiles → IQR

X <ht>

$\langle x_1, x_2, | x_3 \ldots | \ldots x_{10} \rangle$ (ascending order)

part 1 | part 2 | . . . . . part 5

$Q_1 \quad Q_2 \quad Q_3 \quad Q_4$

05 5.1	4.5 5.2	4.8 5.6	4.1 5.9	4.9 5.3

$Q_1 \quad Q_2 \quad Q_3 \quad Q_4$

defn : Quantiles are points taken at regular intervals of a distribution, dividing it into equal size consecutive sets.

NOTE : The 2-quantile is the data point dividing the lower and upper halves of data distribⁿ. This corresponds to median.

Quartiles:

→ line up the data in ascending order.

→ split the data into 04 equally sized chunks. with each chunk containing one quarter of data.

→ lowest quartile is Q1 and highest quartile or upper quartile is Q3.

$$\boxed{\begin{aligned} IQR &= \text{upper quartile} - \text{lower quartile} \\ &= Q_3 - Q_1 \end{aligned}}$$

Eg 1) : < 30 36 $\boxed{47}$ | 50 52 $\boxed{52}$ | 56 60 $\boxed{63}$ | 70 70 110 >

Q_1 Q_2 Q_3

$Q_1 = 47$ $\quad Q_2 = 52$ $\quad Q_3 = 63$

$\frac{n}{4} = \frac{12}{4} = 03$

$Q_3 = \frac{3n}{4} = \frac{3 \times \cancel{12}^{3}}{\cancel{4}} = 09$

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 63 - 47 \\ &= \underline{\underline{16}} \end{aligned}$$

2) Given below are the scores of a player :

points per game	3	6	7	10	11	13	30
f	2	1	2	3	1	1	1

LB (above 3), UB (above 30)

$\Sigma f = 11$

i) Range $= 30 - 3 = 27$

ii) $Q_1 = \frac{n}{4} = \frac{11}{4} = 2.75 \approx 03$

3, 3, $\underline{6}$

Value at $Q_1 = 06$

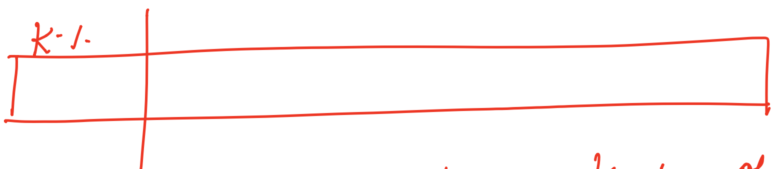
iii) $Q_3 = \frac{3n}{4} = \frac{3 \times 11}{4} = 8.25 \approx 09$

Value at $Q_3 = 11$

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 11 - 06 = \underline{\underline{05}} \end{aligned}$$

→ Instead of splitting data into 04 parts, it may be splitted into 10 parts.. The values are called <u>deciles</u>.

→ If the data is split into percentages, the values that split the data are called <u>percentiles</u>.



P_K is the value K% of the way through the data.

Eg: Assume there are 125 numbers. Find P_{10}.

$$K\left(\frac{n}{100}\right) = 1\cancel{0} \times \frac{125}{10\cancel{0}} = 12.5 \approx 13$$

10 th percentile is the value at posn 13.

Box - plots :

Box and whisker diagram.

<Range, IQR and median>

LB = 03 UB = 30 LQ = 06 (Q1) UQ = 11 (Q3) median = 10

IQR = 5