Thurs	7:15 AM	7 AM
Fri	7 AM	10 AM
Sat	? 7-8	? 9-10

→ Variance are high, prediction are low
less prediction — high Variance

31/10/25

Measure of Dispersion :-

→ Variance
→ Standard Deviation
→ Range.

Variance :-

→ Population Variance (σ^2)

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

→ Sample Variance (s^2)

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

$n-1$ → degree of freedom [Interview Q^n] ***

Ex:- {1, 2, 2, 3, 4, 5}

Variance :- $\frac{17^2}{6} = 2.8$

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{N}$$

$$= \frac{(1-2.8)^2 + (2-2.8)^2 + (2-2.8)^2 + (3-2.8)^2 + (4-2.8)^2 + (5-2.8)^2}{6}$$

$$= \frac{(-1.8)^2 + (-0.8)^2 + (-0.8)^2 + (0.2)^2 + (1.2)^2 + (2.2)^2}{6}$$

$$= \frac{3.24 + 0.64 + 0.64 + 0.04 + 1.44 + 7.84}{6}$$

$$= \frac{13.84}{6}$$

$$\boxed{\sigma^2 = 2.306666}$$

Standard Deviation : $\sqrt{\sigma^2}$

$$\sigma = \sqrt{\sigma^2}$$

$$\sigma = \sqrt{\text{variance}}$$

$$\sigma = \sqrt{2.30} = 1.516$$

To normalize the units

Sample Variance:

$$s^2 = \frac{13.84}{5} = 2.768$$

Standard Deviation $= \sqrt{s^2}$

$$= \sqrt{2.768}$$

$$= 1.66$$

Range: Maximum − Minimum.

Ex:- {1, 2, 2, 3, 4, 5}

Range = 5 − 1

= 4.

Percentile & Quartile:-

Percentile:- It is the value below which a certain Percent of observations will lie (come)

Ex:- {1,1,2,3,4,5,5,6,7,7,8} = 11 [100%]

How much % of data will come below 6?

$$\text{Percentile rank of } x = \frac{\text{No. of value below } x}{N} \times 100$$

$$= \frac{7}{11} \times 100$$

= 63% observation data value is < 6.

Quartile:- It helps to find the value which is present at the given percentile rank.

Ex: {1,1,2,3,4,5,5,6,7,7,8}
(0 1 2 3 4 5 6 7 8 9 10)

which value is present at 25%?

$$\text{Value} = \frac{\text{Percentile}}{100} \times [n+1]$$

$$= \frac{25}{100} \times [12]$$

$$V = 3 \rightarrow [\text{Index}]$$

Value = 2.

90%

$$V = \frac{90}{100} \times 12$$

$$= 10.8$$

→ A percentile indicates the value below which a given percentage of observations fall in a dataset.

∴ There are 99 percentiles ($P_1, P_2 \cdots P_{99}$).

Ex:- 50^{th} percentile (P_{50}) means

50% of data are below that value.

formula (for ungrouped data):-

$$P_K = \left(\frac{K(N+1)}{100}\right)^{th} \text{value}$$

$P_K = K^{th}$ percentile

N = No. of observations.

Ex: 40, 50, 55, 60, 65, 70, 75, 80, 90, 95

$30^{th}\% = ?$

$$P_{30} = \frac{\left[\frac{30}{100}\right]\left[10+1\right]}{100} = 3.3^{rd} \text{ value}$$

$\leftarrow$ between 3 & 4

→ Quartiles divide the data into 4 Equal parts, Each Containing 25% of Observations.

$Q_1 \rightarrow 25^{th}$ Percentile (P_{25}) → 25% of data below it.

$Q_2 \rightarrow 50^{th}$ → 50% of data below it median

$Q_3 \rightarrow 75^{th}$ → 75% of data below it

Ungrouped :- $Q_K = \left(\frac{K(N+1)}{4}\right)^{th}$ value, $K = 1, 2, 3$

Ex: 10, 20, 30, 40, 50, 60, 70, 80

$$Q_1 = \left[\frac{1 \times (8+1)}{4}\right]^{th} = 2.25^{th} \text{ b/w 2 \& 3}$$

Five number Summary:-

1) minimum
2) First Quartile (Q_1) 25% below
2) Median (Q_2) 50%
3) Third Quartile (Q_3) 75%
5) Maximum.

Note:- choose these 5 num after removing the outliers from the data by finding boundary values.

→ **Lower Fence & Upper Fence**

→ $LF = Q_1 - 1.5\,(IQR)$

$UF = Q_3 + 1.5\,(IQR)$

IQR (Inter Quartile Range)

$IQR = (Q_3 - Q_1)$

Ex:- {1, 1, 2, 3, 4, 4, 4, 5, 5, 6, 7, 7, 8, 8, 9, 15, 28, 36} = 17

$LF = 3 - 1.5(3)$

$= -4.5$

$UF = 8 + 1.5(5)$

$= 15.5$

$IQR = Q_3 - Q_1$

$= 8 - 3$

$= 5$

minimum = 1, median = 5, max = 9.

-4.5 & 15.5 is outlier.

$Q_1 = \frac{25}{100} \times 18$

$= 3$

$Q_2 = 5.$

$Q_3 = 8$

$Q_4 = 9.$