Range :

$$X_{max} - X_{min}$$

Interquartile range

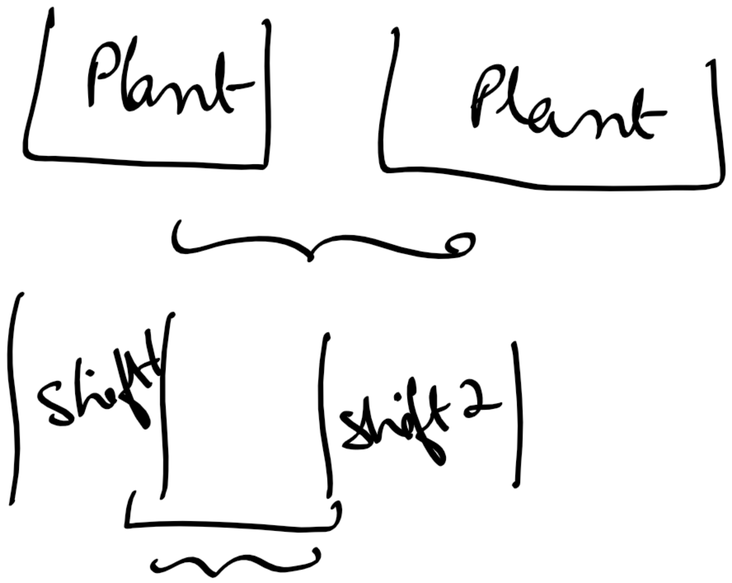
$$Q_3 - Q_1$$

Quartile Deviation (semi interquartile range)

$$\boxed{Q.D = \frac{(Q_3 - Q_1)}{2}}$$

Dr. HARINAKSHI KARKERA

less S.D ⇒ less risk



Dr. HARINAKSHI KARKERA

Root mean square deviation

$$s = \sqrt{\frac{1}{N}\sum_i f_i (x_i - A)^2}$$

A – any arbitrary number

$A = \bar{x}$

$s = \sigma$

Mean square deviation

$$s^2 = \frac{1}{N}\sum_i f_i (x_i - A)^2$$

$A = \bar{x}$

σ^2

Dr. HARINAKSHI KARKERA

σ s

Relation between standard deviation and Root mean square deviation

Relation between σ and s. By definition, we have

$$s^2 = \frac{1}{N}\sum_i f_i (x_i - A)^2 = \frac{1}{N}\sum_i f_i (x_i - \bar{x} + \bar{x} - A)^2$$

$$= \frac{1}{N}\sum_i f_i \left[(x_i - \bar{x})^2 + (\bar{x} - A)^2 + 2(\bar{x} - A)(x_i - \bar{x}) \right]$$

$$= \frac{1}{N}\sum_i f_i (x_i - \bar{x})^2 + (\bar{x} - A)^2 \frac{1}{N}\sum_i f_i + 2(\bar{x} - A)\sum_i f_i (x_i - \bar{x})$$

$$\boxed{s^2 = \sigma^2 + (\bar{x} - A)^2 = \sigma^2 + d^2}, \text{ where } d = \bar{x} - A$$

$d = 0 \Rightarrow \bar{x} = A$ s^2 is least

Coefficient of Dispersion

C. D. based on range = $\frac{A-B}{A+B}$, where A and B are the greatest and the smallest items in the series.

Based upon Quartile deviation : C.D. = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$.

Based upon mean deviation: C.D. = $\frac{Mean\ deviation}{Average\ from\ which\ it\ is\ calculated}$..

Based upon standard deviation : C. D. = $\frac{\sigma}{\bar{x}}$.

$C \cdot D = \frac{\sigma}{\bar{x}}$

Coefficient of Variation

100 times the coefficient of dispersion based upon standard deviation is called coefficient of variation.

C. V. =100×$\frac{\sigma}{\bar{x}}$.

Dr. HARINAKSHI KARKERA

* An analysis of monthly wages paid to the workers of two firms A & B belonging to the same industry gives the following results :

	Firm A	Firm B
No. of workers	500	600
Average daily wage (₹)	186	175
Variance of distⁿ of wages	81	100

(i) Which firm, A or B, has a larger wage bill ?

(ii) In which firm, A or B, is there greater variability in individual wages

(iii) Calculate (a) the average daily wage of all the workers in the firms A & B taken together

(b) the variance of the distⁿ of wages

Dr. HARINAKSHI KARKERA

$n_1 = 500$ $\qquad$ $n_2 = 600$

$\bar{x}_1 = 186$ $\qquad$ $\bar{x}_2 = 175$

$\sigma_1^2 = 81$ $\qquad$ $\sigma_2^2 = 100$

(i) $n_1 \times \bar{x}_1 = 500 \times 186 = 93{,}000$ (₹) 'A'

$n_2 \times \bar{x}_2 = 600 \times 175 = 1{,}05{,}000$ (₹) 'B'

Firm B has larger wage bill

(ii) For firm A, $\sigma_1^2 = 81$, $\sigma_1 = 9$

For firm B, $\sigma_2^2 = 100$, $\sigma_2 = 10$

C.V of distn of wages for firm A $= 100 \times \frac{\sigma_1}{\bar{x}_1}$

$= \frac{100 \times 9}{186} = 4.838$

Dr. HARINAKSHI KARKERA

(ii) For firm A, $\sigma_1^2 = 81$, $\sigma_1 = 9$

For firm B, $\sigma_2^2 = 100$, $\sigma_2 = 10$

C.V of distn of wages for firm A $= 100 \times \frac{\sigma_1}{\bar{x}_1}$

$= \frac{100 \times 9}{186} = \underline{\underline{4.838}}$

C.V of distn of wages for Firm B $= 100 \times \frac{\sigma_2}{\bar{x}_2}$

$= \frac{100 \times 10}{175}$

$= \underline{\underline{5.714}}$

$\therefore$ Firm B has greater variability in individual wages.

Dr. HARINAKSHI KARKERA

(iii) a) Combined mean $\bar{x} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}$

$$= \frac{500 \times 186 + 600 \times 175}{500+600}$$

$$= \underline{\underline{180}} \ (₹)$$

b) Combined variance

$$\sigma^2 = \frac{1}{n_1+n_2}\left[n_1\sigma_1^2 + n_2\sigma_2^2 + \frac{n_1 n_2}{n_1+n_2}(\bar{x}_1 - \bar{x}_2)^2\right]$$

$$= \frac{1}{500+600}\left[500 \times 81 + 600 \times 100 + \left(\frac{500 \times 600}{500+600}\right)(\quad)^2\right]$$

$$\sigma^2 = \underline{\underline{121.36}}$$

Dr. HARINAKSHI KARKERA

Moments

The rth moment of a variable x about any point $x = A$, denoted by μ_r'

$$\mu_r' = \frac{1}{N} \sum_i f_i (x_i - A)^r \qquad N = \sum_i f_i$$

$$\boxed{\mu_r' = \frac{1}{N} \sum f_i d_i^r} \qquad d_i = x_i - A$$

Central moments

The rth moment of a variable x about the mean $\bar{x}$ μ_r

$$z_i = x_i - \bar{x}$$

$$\boxed{\mu_r = \frac{1}{N} \sum_i f_i (x_i - \bar{x})^r = \frac{1}{N} \sum_i f_i z_i^r}$$

DR. HARINAKSHI KARKERA

$$\mu_r = \frac{1}{N}\sum_i f_i (x_i - \bar{x})^r = \frac{1}{N}\sum_i f_i Z_i^r$$

1) Put $r = 0$

$$\mu_0 = \frac{1}{N}\sum f_i (x_i - \bar{x})^0$$

$$= \frac{\sum f_i}{N}$$

$$\boxed{\mu_0 = 1}$$

2) Put $r = 1$

$$\mu_1 = \frac{1}{N}\sum f_i (x_i - \bar{x})$$

$$\boxed{\mu_1 = 0}$$

3) Put $r = 2$

$$\mu_2 = \frac{1}{N}\sum f_i (x_i - \bar{x})^2$$

$$= \sigma^2$$

$$\mu_1' = \frac{1}{N}\sum f_i (x_i - A)$$

$$\boxed{d_i = x_i - A}$$

$$\bar{x} = A + \underbrace{\frac{1}{N}\sum f_i d_i}$$

$$\boxed{\bar{x} = A + \mu_1'} \qquad \boxed{\mu_1' = \bar{x} - A}$$

Dr. HARINAKSHI KARKERA

- **Relationship between raw and central moments:**

$$\mu_1' = \bar{x} - A$$

$$\mu_0 = \mu_0' = 1$$

$$\mu_1 = 0$$

$$\boxed{\mu_2 = \mu_2' - \mu_1'^2}$$

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3$$

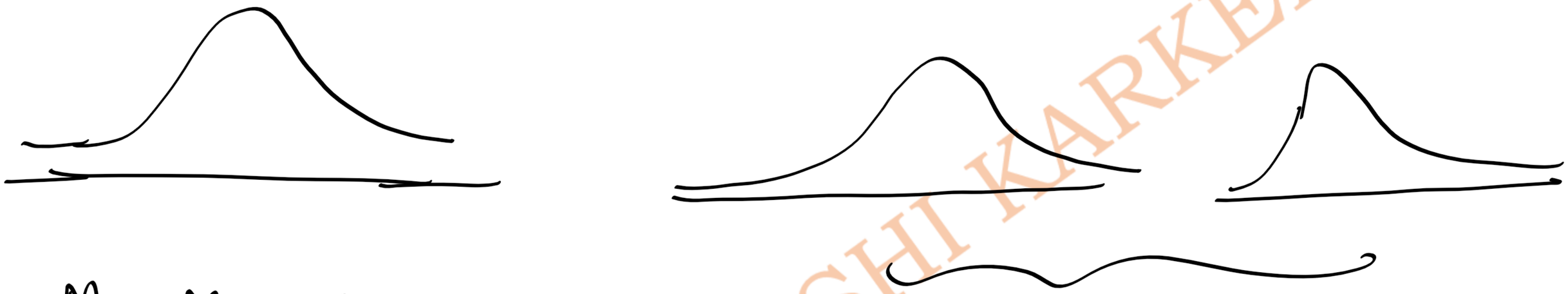
$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$

$$\mu_2' = \mu_2 + \mu_1'^2$$

$$\mu_3' = \mu_3 + 3\mu_2\mu_1' + \mu_1'^3$$

Skewness

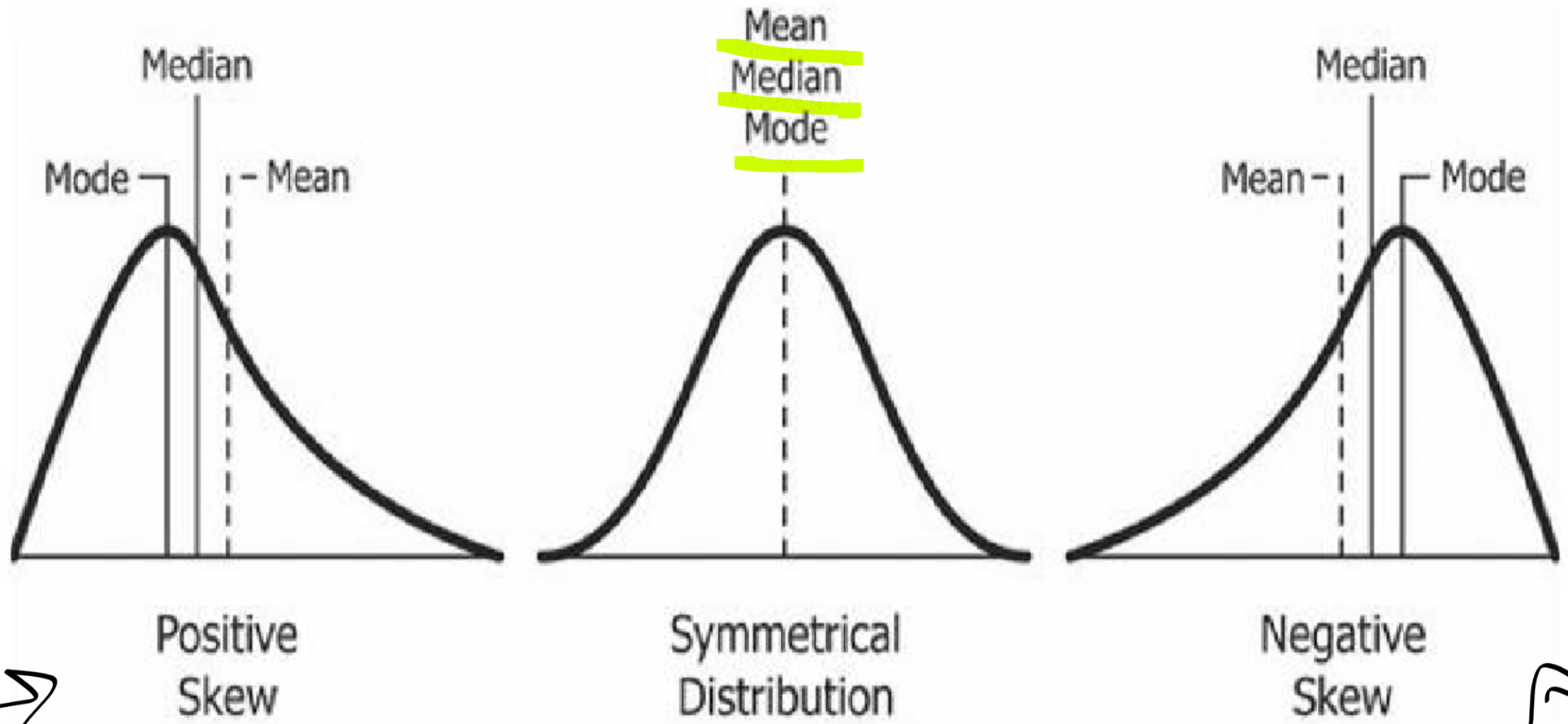
Skewness = lack of symmetry



$M = M_d = M_o$

Dr. HARINAKSHI KARKERA

There are two types of Skewness: Positive and Negative



Mean > Median > Mode

Mean < Median < Mode

Dr.

Measures of skewness

M = mean
M_d = Median
M_o = Mode
of the dist^n.

(1) $S_k = M - M_d$

(2) $S_k = M - M_o$

(3) $S_k = (Q_3 - M_d) - (M_d - Q_1)$

$S_k > 0$ positively skewed (right skewed)

$S_k = 0$ symmetric dist^n

$S_k < 0$ negatively skewed (left skewed)

Dr. HARINAKSHI KARKERA

Coefficient of Skewness

(1) Karl Pearson's coeff. of skewness

$$S_k = \frac{(M - M_0)}{\sigma}$$

σ — S.D of the distn

$$M_0 = 3M_d - 2M$$

$$\boxed{S_k = \frac{3(M - M_d)}{\sigma}}$$

(2) Bowley's coeff. of skewness (Based on quartiles)

$$\boxed{S_k = \frac{Q_3 + Q_1 - 2M_d}{(Q_3 - Q_1)}}$$

Dr. HARINAKSHI KARKERA

- If the skewness is between -0.5 and 0.5, the data are fairly symmetrical.

- If the skewness is between -1 and -0.5(negatively skewed) or between 0.5 and 1(positively skewed), the data are moderately skewed.

- If the skewness is less than -1(negatively skewed) or greater than 1(positively skewed), the data are highly skewed.

Dr. HARINAKS

* Find the second, third and fourth central moments of the frequency distribution given below. Hence find the measure of skewness (γ_1).

Class limits	Frequency
110.0 — 114.9	5
115.0 — 119.9	15
120.0 — 124.9	20
125.0 — 129.9	35
130.0 — 134.9	10
135.0 — 139.9	10
140.0 — 144.9	5

Dr. HARINAKSHI KARKERA

* Age distribution of hundred life insurance policy holders is as follows: Calculate the mean deviation from median age.

Age as on nearest birthday	Number
17 – 19.5	9
20 – 25.5	16
26 – 35.5	12
36 – 40.5	26
41 – 50.5	14
51 – 55.5	12
56 – 60.5	6
61 – 70.5	5
	100

Dr. HARINAKSHI KARKERA