8. MEASURES OF DISPERSION

Sections	No. of periods (6)	Weightage in IPE [1x2 + 1x7 =9]
Measures of Dispersion	6	2 & 7 marks

- Statistics is the science of counting.
- Statistics is the science of averages.
- Statistics is the science of estimates and probabilities.
- Statistics is the science of collection, presentation, analysis and interpretation of numerical data.
- Basic terminology of statistics is introduced in the begining. Ungrouped data discretes frequency distribution and continuous frequency distribution are introduced.
- Illustrative Examples are divided into three parts.
- First part consists of examples on finding Mean deviation about Mean, Mean deviation about Median, Variance and Standard deviation for ungrouped data.
- Second part consists of examples on finding Mean deviation about Mean, Mean deviation about Median, Variance and Standard deviation for discrete frequency distribution.
- Third part consists of examples on finding Mean deviation about Mean, Mean deviation about Median, Variance and Standard deviation for continuous frequency distribution. This part concludes with a few model problems on Coefficient of Variation (C.V).

SYNOPSIS POINTS

In the lower classes, we learnt that Mean, Median, Mode, G.M and H.M are called Measures of central tendency. These measures give a rough idea about where the given data points are centred.

I. MEAN($\bar{x}$) AND MEDIAN (M)

Data type	Mean($\bar{x}$)	Median(M)
a. Ungrouped data	$\bar{x} = \frac{\text{Sum of items}}{\text{No. of items}}$	$M = \left(\frac{n+1}{2}\right)^{th}$ item, for **odd n** M= Average of $\frac{n}{2}, \frac{n+2}{2}$ items, for **even n**
b. Grouped data (Discrete)	$\bar{x} = \frac{\Sigma f_i x_i}{N}$ $N = \Sigma f_i$	M = The xi whose c.f is equal or just greater than $\frac{N}{2}$; $N = \Sigma fi$
c. Grouped data (Continuous)	$\bar{x} = \frac{\Sigma f_i x_i}{N}$ $N = \Sigma f_i$ xi = Midpoints of class intervals	$M = l + \frac{C}{f}\left(\frac{N}{2} - m\right)$ Median class = class of $\frac{N}{2}$ th item l = lower limit of median class C = width of class interval f = frequency of median class. m = cumulative frequency of the class just preceeding the median class

II. Mean from Step deviation Method (Shortcut Method):

$$\bar{x} = A + C\left(\frac{\Sigma f_i d_i}{N}\right)$$

(or) $\bar{x} = A + C\bar{d}$

where, A = Assumed mean

C = width of class interval

$$di = \frac{xi - A}{C}$$

$$\bar{d} = \frac{\Sigma f_i d_i}{N}$$

Mean Deviation, Standard Deviation, Variance and Coefficient of Variation are called Measures of Dispersion. These measures describe the spread of the individual values around the central value.

III. Mean Deviation(M.D):

Data type	About Mean	About Median
a. Ungrouped data	$M.D = \frac{\Sigma\|x_i - \bar{x}\|}{n}$	$M.D = \frac{\Sigma\|x_i - M\|}{n}$
b. Grouped data (Discrete)	$M.D = \frac{\Sigma f_i\|x_i - \bar{x}\|}{N}$ $N = \Sigma f_i ; \bar{x} = \frac{\Sigma f_i x_i}{N}$	$M.D = \frac{\Sigma f_i\|x_i - M\|}{N}$ M = The x_i whose c.f is equal or just greater than $\frac{N}{2}$; $N=\Sigma f_i$
c. Grouped data (Continuous)	$M.D = \frac{\Sigma f_i\|x_i - \bar{x}\|}{N}$ Here $\bar{x}$ is calculated using step deviation method	$M.D = \frac{\Sigma f_i\|x_i - M\|}{N}$ $M = l + \frac{C}{f}\left(\frac{N}{2} - m\right)$

IV. Variance (σ^2) and Standard Deviation (σ):

Data type	Variance (σ^2)	Standard Deviation(σ)
a. Grouped data (Discrete)	$\sigma^2 = \frac{\Sigma(x_i - \bar{x})^2}{n}$ (or) $\sigma^2 = \frac{\Sigma x_i^2}{n} - \bar{x}^2$	$\sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$ $\sigma = \sqrt{\frac{\Sigma x_i^2}{n} - \bar{x}^2}$
b. Grouped data (Continuous)	$\sigma^2 = C^2\left[\frac{\Sigma f_i d_i^2}{N} - \bar{d}^2\right]$ C = Width of the interval $N = \Sigma f_i$ $d_i = \frac{x_i - A}{C}$,A=Assumed mean $\bar{d} = \frac{\Sigma f_i d_i}{N}$	$\sigma = C\sqrt{\left[\frac{\Sigma f_i d_i^2}{N} - \bar{d}^2\right]}$ C = Width of the interval $N = \Sigma f_i$ $d_i = \frac{x_i - A}{C}$,A=Assumed mean $\bar{d} = \frac{\Sigma f_i d_i}{N}$

V. Coefficient of variation C.V = $\frac{S.D}{Mean} \times 100$ (or) $C.V = \frac{\sigma}{\bar{x}} \times 100$

ADDITIONAL QUESTIONS WITH SOLUTIONS

1.1 **The following table gives the daily wages of workers in a factory.Compute the standard deviation and coefficient of variation of the wages of the workers:**

Wages(Rs.)	125-175	175-225	225-275	275-325	325-375	375-425	425-475	475-525	525-575
No. of workers	2	22	19	14	3	4	6	1	1

Sol : Here we take assumed mean A= 300. Here C= 50, then $d_i = \frac{x_i - 300}{50}$

Midpoint of C.I (x_i)	frequency (f_i)	d_i	$f_i d_i$	$f_i d_i^2$
150	2	–3	–6	18
200	22	–2	–44	88
250	19	–1	–19	19
300	14	0	0	0
350	3	1	3	3
400	4	2	8	16
450	6	3	18	54
500	1	4	4	16
550	1	5	5	25
	$\Sigma f_i=72=N$		$\Sigma f_i d_i = -31$	$\Sigma f_i d_i^2 = 239$

Here, N=72, $\Sigma f_i d_i = -31$, $\Sigma f_i d_i^2 = 239$

Mean $\bar{x} = A + C\left(\frac{\Sigma f_i d_i}{N}\right) = 300 + 50\left(\frac{-31}{72}\right) = 300 - \frac{1550}{72} = 278.47$

Variance $\sigma^2 = C^2\left[\frac{\Sigma f_i d_i^2}{N} - \left(\frac{\Sigma f_i d_i}{N}\right)^2\right]$

$$= 50^2\left[\frac{239}{72} - \left(\frac{-31}{72}\right)^2\right] = 2500\left[\frac{239}{72} - \frac{961}{72(72)}\right] = 2500\left[\frac{(239)(72)-(961)}{72(72)}\right]$$

$$= 2500\left[\frac{17208-961}{72(72)}\right] = 2500\left[\frac{16247}{72(72)}\right] = 7835.17$$

$\therefore$ Standard deviation $\sigma = \sqrt{7835.17} = 88.51$

Coefficient of variation $C.V = \frac{\sigma}{\bar{x}} \times 100 = \frac{88.51}{278.47} \times 100 = 31.79$

1.2 The scores of two crickters A and B in 10 innings are given below.Find who is a better run getter and who is a more consistent player.

Scores of A : x_i	40	25	19	80	38	8	67	121	66	76
Scores of B : y_i	28	70	31	0	14	111	66	31	25	4

Sol : For cricketer A, $\Sigma x_i=540$, and n=10. ∴ Mean $\bar{x}=\frac{540}{10}=54$;

For cricketer B, $\Sigma y_i=380$, and n=10. ∴ Mean $\bar{y}=\frac{380}{10}=38$

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	y_i	$y_i-\bar{y}$	$(y_i-\bar{y})^2$
40	–14	196	28	–10	100
25	29	841	70	32	1024
19	–35	1225	31	–7	49
80	26	676	0	–38	1444
38	–16	256	14	–24	576
8	–46	2116	111	73	5329
67	13	169	66	28	784
121	67	4489	31	–7	49
66	12	144	25	–13	169
76	22	484	4	–34	1156
$\Sigma x_i=540$		10596	$\Sigma y_i=380$		10680

For the scores of A, Standard deviation $\sigma_x=\sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n}}=\sqrt{\frac{10596}{10}}=\sqrt{1059.6}=32.55$

For the scores of B, Standard deviation $\sigma_y=\sqrt{\frac{\Sigma(y_i-\bar{y})^2}{n}}=\sqrt{\frac{10680}{10}}=\sqrt{1068}=32.68$

For A, Coefficient of variation C.V $=\frac{\sigma_x}{\bar{x}}\times 100=\frac{32.55}{54}\times 100=60.28$

For B, Coefficient of variation C.V $=\frac{\sigma_y}{\bar{y}}\times 100=\frac{32.68}{38}\times 100=86$

Greater mean represents the better result and Lesser variation represents more stability.
Since $\bar{x}>\bar{y}$, cricketer A is a better run getter (scorer).
Since C.V. of A < C.V. of B, cricketer A is also a more consistent player.

1.3 **From the prices of share X and Y given below for 10 days of trading, find out which share is more stable?**

X	35	54	52	53	56	58	52	50	51	49
Y	108	107	105	105	106	107	104	103	104	101

Sol: Mean of share X is $\bar{x}_1 = \frac{35+54+52+53+56+58+52+50+51+49}{10} = \frac{510}{10} = 51$

Variance of share X is $\sigma_1^2 = \frac{\Sigma x_i^2}{n} - (\bar{x}_1)^2$

$= \frac{1225+2916+2704+2809+3136+3364+2704+2500+2601+2401}{10} - (51)^2$

$= \frac{26360}{10} - 2601 = 2636 - 2601 = 35$

Standard deviation of X is $\sigma_1 = \sqrt{35} = 5.92$

Coefficient of variation of X is C.V$= \frac{\sigma_1}{\bar{x}_1} \times 100 = \frac{5.92}{51} \times 100 = 11.61$ .....(1)

Mean of share Y is

$\bar{x}_2 = \frac{108+107+105+105+106+107+104+103+104+101}{10} = \frac{1050}{10} = 105$

Variance of share Y is $\sigma_2^2 = \frac{\Sigma x_i^2}{n} - (\bar{x}_2)^2$

$= \frac{11664+11449+11025+11025+11236+11449+10816+10609+10816+10201}{10} - (105)^2$

$= \frac{110290}{10} - 11025 = 11029 - 11025 = 4$

Standard deviation of Y is $\sigma_2 = \sqrt{4} = 2$

Coefficient of variation of Y is $\frac{\sigma_2}{\bar{x}_2} \times 100 = \frac{2}{105} \times 100 = 1.91$ .....(2)

Less variation represents more stability.

From (1) & (2), C.V of Y is lower than C.V of X.

$\therefore$ Share of Y is more stable.