MEASURES OF DISPERSION

Measures of Central Tendency and Dispersion

Measure of Central Tendency:

1) Mathematical Average: a) Arithmetic mean (A.M.)

b) Geometric mean (G.M.)

c) Harmonic mean (H.M.)

2) Averages of Position: a) Median

b) Mode

Arithmetic Mean:

(1) Simple arithmetic mean in individual series

(i) Direct method: If the series in this case be $x_1, x_2, x_3, ..., x_n$; then the arithmetic mean $\bar{x}$ is given by

$$\bar{x} = \frac{\text{Sum of the series}}{\text{Number of terms}}$$

$$i.e., \bar{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

(2) Simple arithmetic mean in continuous series If the terms of the given series be $x_1, x_2, ..., x_n$ and the corresponding frequencies be $f_1, f_2, ... f_n$, then the arithmetic mean $\bar{x}$ is given by,

$$\overline{x} = \frac{f_1x_1 + f_2x_2 + \dots + f_nx_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i}.$$

Continuous Series: If the series is continuous then x_{ii}'s are to be replaced by m_i's where m_i's are the mid values of the class intervals.

Mean of the Composite Series: If $\overline{x}_i, (i = 1, 2 \dots, k)$ are the means of k-component series of sizes $n_i, (i = 1, 2, \dots, k)$ respectively, then the mean $\overline{x}$ of the composite series obtained on combining the component series is given by the formula $\overline{x} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2 + \dots + n_k\overline{x}_k}{n_1 + n_2 + \dots + n_k} = \frac{\sum_{i=1}^{n} n_i\overline{x}_i}{\sum_{i=1}^{n} n_i}$.

Geometric Mean: If $x_1, x_2, x_3, \dots, x_n$ are n values of a variate x, none of them being zero, then geometric mean (G.M.) is given by $\text{G.M.} = (x_1.x_2.x_3 \dots x_n)^{1/n}$

In case of frequency distribution, G.M. of n values $x_1, x_2, \dots x_n$ of a variate x occurring with frequency $f_1, f_2, \dots, f_n$ is given by $\text{G.M.} = (x_1^{f_1}.x_2^{f_2} \dots x_n^{f_n})^{1/N}$, where $N = f_1 + f_2 + \dots + f_n$.

Continuous Series: If the series is continuous then x_{ii}'s are to be replaced by m_i's where m_i's are the mid values of the class intervals.

Harmonic Mean: The harmonic mean of n items $x_1, x_2, \dots, x_n$ is defined as $\text{H.M.} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$.

If the frequency distribution is $f_1, f_2, f_3, \dots, f_n$ respectively, then $\text{H.M.} = \frac{f_1 + f_2 + f_3 + \dots + f_n}{\left(\frac{f_1}{x_1} + \frac{f_2}{x_2} + \dots + \frac{f_n}{x_n}\right)}$.

Median: The median is the central value of the set of observations provided all the observations are arranged in the ascending or descending orders. It is generally used, when effect of extreme items is to be kept out.

(1) Calculation of median

(i) **Individual series:** If the data is raw, arrange in ascending or descending order. Let *n* be the number of observations.

If *n* is odd, Median = value of $\left(\frac{n+1}{2}\right)^{th}$ item.

If *n* is even, Median $=\frac{1}{2}\left[\text{value of }\left(\frac{n}{2}\right)^{\text{th}}\text{ item} + \text{value of }\left(\frac{n}{2}+1\right)^{\text{th}}\text{ item}\right]$

(ii) Discrete series: In this case, we first find the cumulative frequencies of the variables arranged in ascending or descending order and the median is given by

Median = $\left(\frac{n+1}{2}\right)^{th}$ observation, where *n* is the cumulative frequency.

(iii) **For grouped or continuous distributions:** In this case, following formula can be used.

(a) For series in ascending order, Median = $l+\frac{\left(\frac{N}{2}-C\right)}{f}\times i$

Where l = Lower limit of the median class

f = Frequency of the median class

N = The sum of all frequencies

i = The width of the median class

C = The cumulative frequency of the class preceding to median class.

(b) For series in descending order

Median = $u-\left(\frac{\frac{N}{2}-C}{f}\right)\times i$, where u = upper limit of the median class, $N=\sum_{i=1}^{n} f_i$.

As median divides a distribution into two equal parts, similarly the quartiles, quintiles, deciles and percentiles divide the distribution respectively into 4, 5, 10 and 100 equal parts. The j^{th} quartile is given by $Q_j = l+\left(\frac{j\frac{N}{4}-C}{f}\right)i; j=1,2,3$. Q_1 is the lower quartile, Q_2 is the median and Q_3 is called the upper quartile.

(2) Lower quartile

(i) Discrete series : $Q_1 = \text{size of} \left(\frac{n+1}{4}\right)^{\text{th}}$ item

(ii) Continuous series : $Q_1 = l + \dfrac{\left(\frac{N}{4} - C\right)}{f} \times i$

(3) Upper quartile

(i) Discrete series : $Q_3 = \text{size of} \left[\frac{3(n+1)}{4}\right]^{\text{th}}$ item

(ii) Continuous series : $Q_3 = l + \dfrac{\left(\frac{3N}{4} - C\right)}{f} \times i$

Mode: The mode or model value of a distribution is that value of the variable for which the frequency is maximum. For continuous series, mode is calculated as,

Mode $= l_1 + \left[\dfrac{f_1 - f_0}{2f_1 - f_0 - f_2}\right] \times i$

Where, l_1 = The lower limit of the model class

f_1 = The frequency of the model class

f_0 = The frequency of the class preceding the model class

f_2 = The frequency of the class succeeding the model class

i = The size of the model class.

Empirical relation : Mean – Mode = 3(Mean – Median) $\Rightarrow$ Mode = 3 Median – 2 Mean.

Measure of dispersion: The degree to which numerical data tend to spread about an average value is called the dispersion of the data. The four measure of dispersion are

(1) Range (2) Mean deviation

(3) Standard deviation (4) Square deviation

(1) Range : It is the difference between the values of extreme items in a series. Range = $X_{max} - X_{min}$

The coefficient of range (scatter) $= \dfrac{x_{max} - x_{min}}{x_{max} + x_{min}}$.

Range is not the measure of central tendency. Range is widely used in statistical series relating to quality control in production.

Range is commonly used measures of dispersion in case of changes in interest rates, exchange rate, share prices and like statistical information. It helps us to determine changes in the qualities of the goods produced in factories.

Quartile deviation or semi inter-quartile range: It is one-half of the difference between the third quartile and first quartile *i.e.*, $\text{Q.D.} = \frac{Q_3 - Q_1}{2}$ and coefficient of quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$, where Q_3 is the third or upper quartile and Q_1 is the lowest or first quartile.

(2) Mean Deviation: The arithmetic average of the deviations (all taking positive) from the mean, median or mode is known as mean deviation.

Mean deviation is used for calculating dispersion of the series relating to economic and social inequalities. Dispersion in the distribution of income and wealth is measured in term of mean deviation.

(i) Mean deviation from ungrouped data (or individual series) Mean deviation $= \frac{\Sigma |x - M|}{n}$,

where $|x - M|$ means the modulus of the deviation of the variate from the mean (mean, median or mode) and n is the number of terms.

(ii) **Mean deviation from continuous series:** Here first of all we find the mean from which deviation is to be taken. Then we find the deviation $dM = |x - M|$ of each variate from the mean M so obtained.

Next we multiply these deviations by the corresponding frequency and find the product $f.dM$ and then the sum $\Sigma f\, dM$ of these products.

Lastly we use the formula, mean deviation $= \frac{\Sigma f |x - M|}{n} = \frac{\Sigma f\, dM}{n}$, where $n = \Sigma f$.

(3) Standard Deviation: Standard deviation (or S.D.) is the square root of the arithmetic mean of the square of deviations of various values from their arithmetic mean and is generally denoted by σ read as sigma. It is used in statistical analysis.

(i) Coefficient of standard deviation: To compare the dispersion of two frequency distributions the relative measure of standard deviation is computed which is known as coefficient of standard deviation and is given by

Coefficient of S.D. $= \frac{\sigma}{\bar{x}}$, where $\bar{x}$ is the A.M.

(ii) Standard deviation from individual series

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

where, $\bar{x}$ = The arithmetic mean of series

N = The total frequency.

(iii) Standard deviation from continuous series

$$\sigma = \sqrt{\frac{\Sigma f_i(x_i-\bar{x})^2}{N}}$$

where, $\bar{x}$ = Arithmetic mean of series

x_i = Mid value of the class

f_i = Frequency of the corresponding x_i

$N = \Sigma f$ = The total frequency

Short cut Method:

(i) $\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$ (ii) $\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$

where, $d = x - A$ = Deviation from the assumed mean A

f = Frequency of the item

$N = \Sigma f$ = Sum of frequencies

(4) Square Deviation:

(i) Root mean square deviation

$$S = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i(x_i - A)^2},$$

where A is any arbitrary number and S is called mean square deviation.

(ii) Relation between S.D. and root mean square deviation : If σ be the standard deviation and S be the root mean square deviation.

Then, $S^2 = \sigma^2 + d^2$.

Obviously, S^2 will be least when $d = 0$ *i.e.*, $\bar{x} = A$

Hence, mean square deviation and consequently root mean square deviation is least, if the deviations are taken from the mean.

Variance: The square of standard deviation is called the variance. Coefficient of standard deviation and variance : The coefficient of standard deviation is the ratio of the S.D. to A.M. *i.e.*, $\frac{\sigma}{\bar{x}}$.

Coefficient of variance = coefficient of S.D.$\times 100 = \frac{\sigma}{\bar{x}} \times 100$.

Variance of the combined series : If n_1, n_2 are the sizes, $\bar{x}_1, \bar{x}_2$ the means and σ_1, σ_2 the standard deviation of two series, then $\sigma^2 = \frac{1}{n_1 + n_2}[n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)]$,

Where $d_1 = \bar{x}_1 - \bar{x}$, $d_2 = \bar{x}_2 - \bar{x}$ and $\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

Very Short Answer Questions

1. Find the mean deviation about the mean for the following data:

i) 38, 70, 48, 40, 42, 55, 63, 46, 54, 44

ii) 3, 6, 10, 4, 9, 10

Sol.i) Mean $\bar{x} = \frac{38+70+48+40+42+55+63+46+54+44}{10}$

$$= \frac{500}{10} = 50$$

The absolute values of mean deviations are $|x_i - \bar{x}|$ = 12, 20, 2, 10, 8, 5, 13, 4, 4, 6.

$\therefore$ Mean deviation about the Mean = $\frac{\sum_{i=1}^{10} |x_i - \bar{x}|}{10}$

$$= \frac{12+20+2+10+8+5+13+4+4+6}{10}$$

$$= \frac{84}{10} = 8.4$$

ii) Mean $(\bar{x}) = \frac{\sum_{i=1}^{6} x_i}{n}$

$$\therefore \bar{x} = \frac{3+6+10+4+9+10}{6} = \frac{42}{6} = 7$$

The absolute values of the deviations are $|x_i - \bar{x}|$ = 4, 1, 3, 3, 2, 3

Mean deviation about the Mean = $\frac{\sum_{i=1}^{6} |x_i - \bar{x}|}{6}$

$$= \frac{4+1+3+3+2+3}{6} = \frac{16}{6} = 2.6666 \simeq 2.67$$

2. Find the mean deviation about the median for the following data.

i) 13, 17, 16, 11, 13, 10, 16, 11, 18, 12, 17

ii) 4, 6, 9, 3, 10, 13, 2

Sol. Given data in the ascending order :10, 11, 11, 12, 13, 13, 16, 16, 17, 17, 18

Mean (M) of these 11 observations is 13.

The absolute values of deviations are $|x_i - M| = 3, 2, 2, 1, 0, 0, 3, 3, 4, 4, 5$

$$\therefore \text{Mean deviation about Median} = \frac{\sum_{i=1}^{11} |x_i - M|}{n} = \frac{3+2+2+1+0+0+3+3+4+4+5}{11}$$

$$= \frac{27}{11} = 2.45$$

ii) 4, 6, 9, 3, 10, 13, 2

Expressing the given data in the ascending order, we get 2, 3, 4, 6, 9, 10, 13.

Median (M) of given data = 6

The absolute values of the deviations are $|x_i - \bar{x}|$ = 4, 3, 2, 0, 3, 4, 7

$$\therefore \text{Mean Deviation about Median} = \frac{\sum_{i=1}^{7} |x_i - M|}{n} = \frac{4+3+2+0+3+4+7}{7} = \frac{23}{7} = 3.29.$$

3. Find the mean deviation about the mean for the following distribution.

i)

x_i	10	11	12	13
f_i	3	12	18	12

ii)

x_i	10	30	50	70	90
f_i	4	24	28	16	8

Sol. i)

x_i	f_i	$f_i x_i$	$\|x_i - \bar{x}\|$	$f_i \|x_i - \bar{x}\|$
10	3	30	1.87	5.61
11	12	132	0.87	10.44
12	18	216	0.13	2.24
13	12	156	1.13	13.56
	N = 45	$\Sigma f_i x_i = 534$		$\Sigma f_i \|x_i - \bar{x}\| = 31.95$

$\therefore$ Mean $(\bar{x}) = \frac{\Sigma f_i x_i}{N} = \frac{534}{45} = 11.87$

$\therefore$ Mean Deviation about the Mean = $\frac{\sum_{i=1}^{4} f_i |x_i - \bar{x}|}{N} = \frac{31.95}{45} = 0.71$.

ii)

x_i	f_i	$f_i x_i$	$\|x_i - \bar{x}\|$	$f_i \|x_i - \bar{x}\|$
10	4	40	40	160
30	24	720	20	480
50	28	1400	0	0
70	16	1120	20	320
90	8	720	40	320
	N = 80	$\Sigma f_i x_i = 4000$		$\Sigma f_i \|x_i - \bar{x}\| = 1280$

$\therefore$ Mean $(\bar{x}) = \frac{\Sigma f_i x_i}{N} = \frac{4000}{80} = 50$

$\therefore$ Mean Deviation about the Mean = $\frac{\sum_{i=1}^{5} f_i |x_i - \bar{x}|}{N} = \frac{1280}{80} = 16$.

4. Find the mean deviation about the median for following frequency distribution.

x_i	5	7	9	10	12	15
f_i	8	6	2	2	2	6

Sol. Writing the observations in ascending order.

x_i	f_i	Cumulative frequency CF)	$\lvert x_i - M \rvert$	$f_i \lvert x_i - M \rvert$
5	8	8	2	16
$7 \rightarrow M$	6	14 > N/2	0	0
9	2	16	2	4
10	2	18	3	6
12	2	20	5	10
15	6	26	8	48
	N = 26			$\Sigma f_i \lvert x_i - M \rvert = 84$

Hence N = 26 and $\frac{N}{2} = 13$

Median (M) = 7

Mean Deviation about Median = $\dfrac{\sum_{i=1}^{6} \lvert x_i - M \rvert}{n} = \dfrac{87}{26} = 3.23$.

Short Answer Questions

1. Find the mean deviation about the median for the following continuous distribution.

i)

Marks obtained	0-10	10-20	20-30	30-40	40-50	50-60
No.of boys	6	8	14	16	4	2

ii)

Class interval	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	5	8	7	12	28	20	10	10

Sol.i)

Class interval	frequency f_i	C.F.	Midpoint x_i	$\lvert x_i - M\rvert$	$f_i \lvert x_i - M \rvert$
0-10	6	6	5	20.86	137.16
10-20	8	14	15	12.86	102.88
20-30	14	28	25	2.86	40.04
30-40	16	44	35	7.14	114.24
40-50	4	48	45	17.14	68.56
50-60	2	50	55	27.14	54.28
	N = 50				517.16

Hence L = 20, $\frac{N}{2} = 25$, $f_1 = 14$, f = 14, h = 10

Median (M) = $L + \left[\frac{\left(\frac{N}{2} - f_i\right)}{f}\right] h = 20 + \frac{25-14}{14} \times 10 = 20 + \frac{110}{14} = 20 + 7.86 = 27.86$

$\therefore$ Mean Deviation about Median = $\frac{\sum_{i=1}^{6} f_i \lvert x_i - M \rvert}{N} = \frac{517.16}{50} = 10.34$.

ii)

Class interval	frequency f_i	C.F.	Midpoint x_i	$\|x_i - M\|$	$f_i \|x_i - M\|$
0-10	5	5	5	41.43	207.15
10-20	8	13	15	31.43	251.44
20-30	7	20	25	21.43	150.01
30-40	12	32	35	11.43	137.16
40-50	28	60	45	1.43	40.04
50-60	20	80	55	8.57	171.40
60-70	10	90	65	18.57	185.70
70-80	10	100	75	28.57	285.70
	N=100				1428.6

Here N = 100, $\frac{N}{2} = 50$, L = 40, f_1 = 32, f = 28, h = 10

Median (M) = $L + \left[\frac{\left(\frac{N}{2} - f_i\right)}{f}\right] h = 40 + \frac{50-32}{28} \times 10 = 40 + \frac{180}{28} = 40 + 6.43 = 46.43$

$\therefore$ Mean Deviation about Median = $\frac{\sum_{i=1}^{8} f_i \,|x_i - M|}{N} = \frac{1428.6}{100} = 14.29$.

2. Find the mean deviation about the mean for the following continuous distribution.

Height (in cms)	95-105	105-115	115-125	125-135	135-145	145-155
Number of boys	9	13	26	30	12	10

Sol.

Height (C.I)	**No.of boys (f_i)**	**Midpoint x_i**	$d_i = \frac{x_i - A}{h}$	**$f_i d_i$**	**$\|x_i - \bar{x}\|$**	**$f_i \|x_i - \bar{x}\|$**
95-105	9	100	–3	–27	25.3	227.7
105-115	13	110	–2	–26	15.3	198.9
115-125	26	120	–1	–26	5.3	137.8
125-135	30	130 → (A)	0	0	4.7	141.0
135-145	12	140	1	12	14.7	176.4
145-155	10	150	2	20	24.7	247.0
	N=100			$\Sigma f_i d_i = -47$		1128.8

Mean $(\bar{x}) = A + \frac{\Sigma f_i d_i}{N} \cdot h = 130 + \left(\frac{-47}{100}\right) \cdot 10 = 130 - 4.7 = 125.3$

$\therefore$ Mean Deviation about Mean = $\frac{\sum_{i=1}^{6} f_i |x_i - \bar{x}|}{N} = \frac{1128.8}{100} = 11.29$.

3. Find the variance for the discrete data given below.

i) 6, 7, 10, 12, 13, 4, 8, 12

ii) 350, 361, 370, 373, 376, 379, 38, 387, 394, 395

Sol.i) Mean $\bar{x} = \frac{6+7+10+12+13+4+8+12}{8} = \frac{72}{8} = 9$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
6	–3	9
7	–2	4
10	1	1
12	3	9
13	4	16
4	–5	25
8	–1	1
12	3	9
		$\Sigma \lvert x_i - \bar{x} \rvert^2 = 74$

Variance $(\sigma^2) = \frac{\sum_{i=1}^{8}(x_i - \bar{x})^2}{n} = \frac{74}{8} = 9.25$.

ii) 350, 361, 370, 373, 376, 379, 38, 387, 394, 395

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
350	–27	729
361	–16	256
370	–7	49
373	–4	16
376	–1	1
379	2	4
385	8	64
387	10	100
394	17	289
395	18	324
		1832

$$\text{Mean } (\bar{x}) = \frac{350+361+370+373+376+379+385+387+394+395}{10} = \frac{3770}{10} = 377$$

$$\text{Variance } (\sigma^2) = \frac{\sum_{i=1}^{10}(x_i - \bar{x})^2}{n} = \frac{1832}{10} = 183.2.$$

4. Find the variance and standard deviation of the following frequency distribution.

x_i	6	10	14	18	24	28	30
f_i	2	4	7	12	8	4	3

Sol.

x_i	f_i	$f_i x_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^2$
6	2	12	–13	169	338
10	4	40	–9	81	324
14	7	98	–5	25	175
18	12	216	–1	1	12
24	8	192	5	25	200
28	4	112	9	81	324
30	3	90	11	121	363
	N=40	760			1736

Mean $(\bar{x}) = \frac{760}{40} = 19$

Variance $(\sigma^2) = \frac{\Sigma f_i (x_i - \bar{x})^2}{N} = \frac{1736}{40} = 43.4$

Standard deviation (σ) = $\sqrt{43.4} = 6.59$.

5. **Find the mean deviation from the mean of the following data, using the step deviation method.**

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No.of students	6	5	8	15	7	6	3

Sol. To find the required statistic, we shall construct the following table.

Class interval	Midpoint x_i	f_i	$d_i=(x_i-35)/10$	$f_i\,d_i$	$\lvert x_i\text{-median}\rvert$	$f_i\,\lvert x_i\text{-median}\rvert$
0-10	5	6	–3	–18	28.4	170.4
10-20	15	5	–2	–10	18.4	92
20-30	25	8	–1	–8	8.4	67.2
30-40	35	15	0	0	1.6	24
40-50	45	7	1	7	11.6	81.2
50-60	55	6	2	12	21.6	129.6
60-70	65	3	3	9	31.6	94.8
		N=50		$\Sigma f_i d_i = -8$		659.2

Here N = 5, $\text{Mean}(\bar{x}) = A + \dfrac{h(\Sigma f_i d_i)}{N}$

$$= 35 + \frac{10(-8)}{50} = 33.4 \text{ marks}$$

Mean Deviation from mean = $\dfrac{1}{N}\Sigma f_i \mid x_i - \bar{x}_i \mid = \dfrac{1}{50}(659.2) = 13.18$ (nearly)

Long Answer Questions

1. Find the mean and variance using the step deviation method of the following tabular data, giving the age distribution of 542 members.

Age in years (x_i)	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No.of members (f_i)	3	61	132	153	140	51	2

Sol.

Age in years C.I.	Mid point (x_i)	(f_i)	$d_i = \frac{x_i - A}{C}$ A = 55, C = 10	$f_i d_i$	d_i^2	$f_i d_i^2$
20-30	25	3	–3	–9	9	27
30-40	35	61	–2	–122	4	244
40-50	45	132	–1	–132	1	132
50-60	55→A	153	0	0	0	0
60-70	65	140	1	140	1	140
70-80	75	51	2	102	4	204
80-90	85	2	3	6	9	18
	N=542			–15	28	765

$$\text{Mean } (\bar{x}) = A + \frac{\Sigma f_i d_i}{N} \times C = 55 + \frac{-15}{542} \times 10 = 55 - 0.277 = 54.723$$

$$\text{Variance } (\mu) = \frac{\Sigma f_i d_i^2}{N} - \left(\frac{\Sigma f_i d_i}{N}\right)^2 = \frac{765}{542} - \left(\frac{-15}{542}\right)^2 = \frac{765}{542} - \frac{225}{(542)^2}$$

$$= \frac{542 \times 765 - 225}{(542)^2} = \frac{414630 - 225}{293764} = \frac{414405}{293764} = 1.4106$$

$$V(\mu) = V\left(\frac{X-A}{C}\right) = \left(\frac{1}{C}\right)^2 \cdot V(X) \qquad \left[\because V(ax+n) = a^2 \cdot V(x)\right]$$

$$V(X) = C^2 \cdot V(\mu) = 100 \times 1.4106 = 141.06.$$

2. The coefficient of variation of two distributions are 60 and 70 and their standard deviations are 21 and 16 respectively. Find their arithmetic means.

Sol. Coefficient of variation (C.V) = $\frac{\sigma}{\bar{x}} \times 100$

i) $60 = \frac{21}{\bar{x}} \times 100 \Rightarrow \bar{x} = 35$

ii) $70 = \frac{16}{\bar{y}} \times 100 \Rightarrow \bar{y} = 22.85$

3. From the prices of shares 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. Variance is independent of charge of origin.

X	**Y**	**X_i^2**	**Y_i^2**
–15	8	225	64
4	7	16	49
2	5	4	25
3	5	9	25
6	6	36	36
8	7	64	49
2	4	4	16

0	3	0	9
1	4	1	16
–1	1	1	1
$\Sigma X_i = 10$	$\Sigma Y_i = 50$	$\Sigma X_i^2 = 360$	$\Sigma Y_i^2 = 290$

$$V(X) = \frac{\Sigma X_i^2}{n} - (\bar{X})^2 = \frac{360}{10} - \left(\frac{10}{10}\right)^2 = 36 - 1 = 35$$

$$V(Y) = \frac{\Sigma Y_i^2}{n} - (\bar{Y})^2 = \frac{290}{10} - \left(\frac{50}{10}\right)^2 = 29 - 25 = 4$$

Y is stable.

4. The mean of 5 observations is 4.4. Their variance is 8.24. If three of the observations are 1, 2 and 6. Find the other two observations.

Sol.

x_i	x_i^2
1	1
2	4
6	36
x	x^2
y	y^2

$$S.D. = \sqrt{\frac{\Sigma m^2}{n} - (\bar{x})^2}$$

$\bar{x} = 4.4$

$$\Rightarrow 4.4 = \frac{1+2+6+x+y}{5}$$

$\Rightarrow 9 + x + y = 22$

$\Rightarrow x + y = 13 \quad ...(1)$

$$S.D.^2 = \frac{1+4+3+x^2+y^2}{5} - (4.4)^2 = \frac{41+x^2+y^2}{5} - 19.36$$

$S.D.^2$ = Variance

$$\text{Variance} = \frac{41+x^2+y^2}{5} - 19.36$$

$$8.24 + 19.36 = \frac{41+x^2+y^2}{5}$$

$$41 + x^2 + y^2 = 5 \times 27.6$$

$$x^2 + y^2 = 138 - 41$$

$$x^2 + y^2 = 97 \quad ...(2)$$

From (1) and (2),

$$x^2 + (13 - x)^2 = 97$$

$$x^2 + 169 + x^2 - 26x = 97$$

$$2x^2 - 26x + 72 = 0$$

$$x^2 - 13x + 36 = 0$$

$$x - 9x - 4x + 36 = 0$$

$$x(x-9) - 4(x-9) = 0$$

$$(x-9)(x-4) = 0$$

$$x = 4, 9$$

Put x = 4 in (1)

$y = 13 - 4 = 9$

Put x = 9 in (1)

$y = 13 - 9 = 4$

$\therefore$ If $x = 4$, then $y = 9$.

If $x = 9$ then $y = 4$.

5. **The arithmetic mean and standard deviation of a set of 9 items are 43 and 5 respectively. If an item of value 63 is added to that set, find the new mean and standard deviation of 10 item set given.**

Sol. $\bar{X} = \dfrac{\sum_{i=1}^{9} x_i}{n}$

$$43 = \frac{\sum_{i=1}^{9} x_i}{9}$$

$$\sum_{i=1}^{9} x_i = 43 \times 9 = 387$$

$$\text{New Mean} = \frac{\sum_{i=1}^{10} x_i}{n} = \frac{\sum_{i=1}^{9} x_i + x_{10}}{10} = \frac{387+63}{10} = 45$$

$$S.D^2 = \frac{\sum_{i=1}^{9} {x_i}^2}{9} - (\bar{x})^2 \Rightarrow 5^2 = \frac{\sum_{i=1}^{9} {x_i}^2}{9} - (43)^2$$

$$\frac{\sum_{i=1}^{9} {x_i}^2}{9} = 25 + 1849 \Rightarrow \frac{\sum_{i=1}^{9} {x_i}^2}{9} = 1874$$

$$\sum_{i=1}^{9} {x_i}^2 = 1874 \times 9 = 16866$$

$$\sum_{i=1}^{10} {x_i}^2 = \sum_{i=1}^{9} {x_i}^2 + x_{10}^2 = 16866 + 3969 = 20835$$

$$\text{New S.D.} = \sqrt{\frac{\sum_{i=1}^{10} {x_i}^2}{10} - (\bar{x})^2} = \sqrt{\frac{20835}{10} - (45)^2}$$

$$= \sqrt{2083.5 - 2025} = \sqrt{58.5} = 7.6485 .$$

6. The following table gives the daily wages of workers in a factor. Compute the standard deviation and the 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
Number of workers	2	22	19	14	3	4	6	1	1

Sol. We shall solve this problem using the step deviation method, since the mid points of the class intervals are numerically large.

Here h = 50. Take a = 300. Then $y_i = \frac{x_i - 300}{50}$

Midpoint x_i	Frequency f_i	y_i	$f_i y_i$	$f_i y_i^2$
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
	N = 72		$\Sigma f_i y_i = -31$	$\Sigma f_i y_i^2 = 239$

Mean $\bar{x} = A + \left(\frac{\Sigma f_i y_i}{M}\right) \times h = 300 + \left(\frac{-31}{72}\right) 50 = 300 - \frac{1550}{72} = 278.47$

Variance $(\sigma_x^2) = \frac{h^2}{N^2}\left[N\Sigma f_i y_i^2 - (\Sigma f_i y_i)^2\right]$

$$= \frac{2500}{72 \times 72}[72(239) - (31 \times 31)]$$

$$\sigma_x = \sqrt{2500\left(\frac{239}{72} - \frac{961}{72\times 72}\right)} = 88.52$$

Coefficient of variation = $\frac{88.52}{278.47}\times 100 = 31.79$.

7. An analysis of monthly wages paid to the workers of two firms A and B belonging to the same industry gives the following data.

	Firm A	Firm B
Number of workers	500	600
Average daily wage (Rs.)	186	175
Variance of distribution of waves	81	100

(i) Which firm A or B, has greater variability in individual wages?

(ii) Which firm has larger wage bill?

Sol. (i) Since variance of distribution of wages in firm A is 81, σ_1^2 = 81 and hence σ_1 = 9.

Since variance of distribution of wages in firm B is 100, σ_2^2 = 100 and hence σ_2 = 10.

$\therefore$ C.V. of distribution of wages of firm A = $\frac{\sigma_1}{x_1}\times 100 = \frac{9}{186}\times 100 = 4.84$

C.V. of distribution of wages of firm B = $\frac{\sigma_2}{x_2}\times 100 = \frac{10}{175}\times 100 = 5.71$

Since C.V. of firm B is greater than C.V. of firm A, we can say that firm B has greater variability in individual wages.

(ii) Firm A has number of workers i.e., wage earners (n_1) = 500.

Its average daily wage, say $\bar{x}_1$ = Rs.186

Since Average daily wage = $\frac{\text{Total wages paid}}{\text{No.of workers}}$, if follows that total wages paid to the workers

$$= n_1\bar{x}_1 = 500\times 186 = \text{Rs.}93,000$$

Firm B has number of wage earners (n_2) = 600

Average daily wage, say $\bar{x}_2 = \text{Rs.}175$

Total daily wages paid to the workers = $n_2\bar{x}_2 = 600 \times 175 = \text{Rs.}1,05,000$.

Hence we see that firm B has larger wage bill.

8. The variance of 20 observations is 5. If each of the observations is multiplied by 2, find the variance of the resulting observations.

Sol. Let the given observations be $x_1, x_2, ..., x_{20}$ and $\bar{x}$ be their mean.

Given that n = 20 and variance = 5

i.e., $\frac{1}{20}\sum_{i=1}^{20}(x_i - \bar{x})^2 = 5$ or $\sum_{i=1}^{20}(x_i - \bar{x})^2 = 100$ ... (1)

If each observation is multiplied by 2, then the new observations are

$y_i = 2x_i,\ i = 1, 2, ..., 20$ or $x_i = y_i / 2$

Therefore $\bar{y} = \frac{1}{20}\sum_{i=1}^{20} y_i = \frac{1}{20}\sum_{i=1}^{20} 2x_i = 2 \cdot \frac{1}{20}\sum_{i=1}^{20} x_i = 2\bar{x}$ or $\bar{x} = \frac{1}{2}\bar{y}$

Substituting the values of x_i and $\bar{x}$ in (1) we get

$$\sum_{i=1}^{20}\left(\frac{1}{2}y_i - \frac{1}{2}\bar{y}\right)^2 = 100 \text{ i.e., } \sum_{i=1}^{20}(y_i - \bar{y})^2 = 400$$

Then the variance of the resulting observations = $\frac{1}{20} \times 400 = 20 = 2^2 \times 5$.

Note: From this example we note that, if each observation in a data is multiplied by a constant k, then the variance of the resulting observations is k^2 time that of the variance of original observations.

9. If each of the observations $x_1, x_2, ..., x_n$ is increased by k, where k is a positive or negative number, then show that the variance remains unchanged.

Sol. Let $\bar{x}$ be the mean of $x_1, x_2, ..., x_n$

Then their variance is given by $\sigma_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

If to each observation we add a constant k, then the new (changed) observations will be

$y_i = x_i + k$ …(1)

Then the mean of the new observations $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i + k) = \frac{1}{n}\left[\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} k\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}(nk) = \bar{x} + k \qquad \text{…(2)}$$

The variance of the new observations = $\sigma_2^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i + k - \bar{x} - k)^2 \text{, using (1) and (2)}$$

$$= \frac{1}{n}(x_i - \bar{x})^2 = \sigma_1^2.$$

Thus the variance of the new observations is the same as that of the original observations.

Note: We note that adding (or subtracting) a positive number to (or form) each of the given set of observations does not affect the variance.

10. The scores of two cricketers 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: $\bar{x} = \frac{540}{10} = 54$

For cricketer B: $\bar{y} = \frac{380}{10} = 38$

x_i	(x_i-median)	(x_i-median)^2	y_i	(y_i- y median)	(y_i - y median)^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

Standard deviation of scores of A = $\sigma_x = \sqrt{\frac{1}{n}\Sigma(x_i - \bar{x})^2} = \sqrt{\frac{10596}{10}} = \sqrt{1059.6} = 32.55$

Standard deviation of scores of B = $\sigma_y = \sqrt{\frac{1}{n}\Sigma(y_i - \bar{y})^2} = \sqrt{\frac{10680}{10}} = \sqrt{1068} = 32.68$

C.V. of A = $\frac{\sigma_x}{\overline{x}} \times 100 = \frac{32.55}{54} \times 100 = 60.28$

C.V. of B = $\frac{\sigma_y}{\overline{y}} \times 100 = \frac{32.68}{38} \times 100 = 86$

Since $\overline{x} > \overline{y}$, 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.