10. Statistics (Measures of Dispersion)

- Range is the difference between the highest and lowest observations of the data.

  Example: The runs scored by a batsman in 6 matches are as follows.
  24, 126, 78, 43, 69, 86
  What is the range of the scores?
  Solution:
  Highest score of the batsman = 126
  Lowest score of the batsman = 24
  ∴ Range = Highest score – Lowest score = 126 – 24 = 102

Quartile Deviation

Quartile deviation is the half of the difference between third quartile, Q_3 and first quartile, Q_1 of the series.

∴ Quartile deviation = Q3-Q12

Quartile deviation gives half of the range of middle 50% observations. Quartile deviation is also known as semi-inter quartile range.

Calculation of Quartile Deviation

1. For an individual series, the first and third quartiles can be calculated using the following formula:

Q_1 = Value of n+14th ordered observation

Q_3 = Value of 3n+14th ordered observation

2. For a discrete series, the first and third quartiles can be calculated using the following formula:

If N = ∑f, then

Q_1 = Value of N+14th ordered observation

Q_3 = Value of 3N+14th ordered observation

3. For a continuos series, the first and third quartiles can be calculated using the following formula:

Q1=L+N4-c.f.f×h

Q3=L+3N4-c.f.f×h

Here, L = lower limit of the quartile class

f = frequency of the quartile class

h = class interval of quartile class

c.f. = total of all the frequencies below the quartile class

N = total frequency, ∑f

- Mean deviation about mean $[\mathrm{M.D.}(\bar{x})]$:
  - **For ungrouped data:** $\mathrm{M.D.}(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$, where $\bar{x}$ is the mean given by $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

- Mean deviation about mean $[\mathrm{M.D.}(\bar{x})]$:
  - **For grouped data:** $\mathrm{M.D.}(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} f_i|x_i - \bar{x}|$, where $\bar{x}$ is the mean given by $\bar{x} = \frac{1}{\mathrm{N}}\sum_{i=1}^{n} f_i x_i$ and

  Example:
  Calculate mean deviation about mean for the following data:

Class	0 – 10	10 – 20	20 – 30	30 – 40

Frequency	21	19	49	11

Solution:
Here, assumed mean (a) = 25 and class size (h) = 10

Class	Frequency (f_i)	Mid-point (x_i)	$d_i = \frac{x_i - 25}{10}$	$f_i d_i$	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
0 – 10	21	5	–2	–42	15	315
10 – 20	19	15	–1	–19	5	95
20 – 30	49	25	0	0	5	245
30 – 40	11	35	1	11	15	165
Total	**100**			**–50**		**820**

Here, $N = \sum_{i=1}^{4} f_i = 100$

Now,

Mean, $\bar{x} = a + \frac{1}{N}\sum_{i=1}^{4} f_i d_i \times h = 25 + \frac{(-50)}{100} \times 10 = 25 - 5 = 20$

$\therefore \text{M.D}\left(\bar{x}\right) = \frac{1}{N}\sum_{i=1}^{4} f_i \left|x - \bar{x}\right| = \frac{1}{100} \times 820 = 8.2$

- **Example:**

Find the mean deviation about the median for the following data:
181, 29, 150, 270, 160, 16, 27, 180, 200

- **Solution:**

Here, the number of observations is 9 and these can be arranged in ascending order as
16, 27, 29, 150, 160, 180, 181, 200, 270

Median, M = $\left(\frac{9+1}{2}\right)^{\text{th}}$ observation or 5^{th} observation = 160

$$\therefore \text{M.D.}(\text{M}) = \frac{1}{9}\sum_{i=1}^{9}\left|x_i - \text{M}\right|$$

$$= \frac{1}{9}\left(|16-160| + |27-160| + |29-160| + |150-160| + |160-160| + |180-160| + |181-160| + |200-160| + |270-160|\right)$$

$$= \frac{1}{9}(144 + 133 + 131 + 10 + 0 + 20 + 21 + 40 + 110)$$

$$= \frac{1}{9} \times 609$$

$$= 67.67$$

- The mean of the squares of the deviations from mean is called the variance and it is denoted by s^2.
- Variance of Data:
  - **For ungrouped data:** $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ (In direct method) or $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2$ (In shortcut method), where $\bar{x}$ is the mean.
- Standard deviation is the square root of variance and it is denoted by s. This means:

$$\text{Standard Deviation} = \sqrt{\text{Variance}}$$

Example:
The mean and standard deviations of 50 observations were calculated as 30 and 4 respectively. Later, it was found that by mistake, 13 was taken instead of 18 for one observation during the calculation. Find the correct mean and the correct standard deviation.

Solution:
It is given that, number of observations (n) = 50
Incorrect mean, $\bar{x} = 30$
Incorrect standard deviation (s) = 4
We know that $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

i.e. $30 = \frac{1}{50}\sum_{i=1}^{50} x_i$ or $\sum_{i=1}^{50} x_i = 1500$

Incorrect sum of observations = 1500

$\therefore$Correct sum of observations = 1500 – 13 + 18 = 1505

$\therefore$ Correct mean = $\frac{1505}{50} = 30.1$

Now, standard deviation, $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2}$

$\Rightarrow \sigma = 4 = \sqrt{\frac{1}{50} \times \text{Incorrect } \sum_{i=1}^{n} x_i^2 - 900}$

$\Rightarrow 16 = \frac{1}{50} \times \text{Incorrect } \sum_{i=1}^{n} x_i^2 - (30)2$

$\Rightarrow \text{Incorrect} \sum_{i=1}^{n} x_i^2 = 916 \times 50 = 45800$

Now, correct $\sum_{i=1}^{n} x_i^2$ = Incorrect $\sum_{i=1}^{n} x_i^2 - (13)^2 + (18)^2 = 45800 - 169 + 324 = 45955$

Correct standard deviation $= \sqrt{\frac{\text{Correct } \Sigma x_i^2}{n} - (\text{Correct mean})^2}$

$= \sqrt{\frac{45955}{50} - (30.1)^2}$

$= \sqrt{919.1 - 906.01}$

$= \sqrt{13.09} = 3.62$

- Variance of Data:

  For discrete frequency distribution: $\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^2$ (In direct method)

  or $\sigma^2 = \frac{1}{N^2}\left[N\sum_{i=1}^{n} f_i x_i^2 - \left(\sum_{i=1}^{n} f_i x_i\right)^2\right]$ (In shortcut method), where $\bar{x}$ is the mean and $N = \sum_{i=1}^{n} f_i$

- Standard deviation is the square root of variance and it is denoted by σ.

  $\text{Standard Deviation} = \sqrt{\text{Variance}}$

- Variance of Data:
  - **For continuous frequency distribution:** $\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^2$ or $\sigma^2 = \frac{1}{N^2}\left[N\sum_{i=1}^{n} f_i x_i^2 - \left(\sum_{i=1}^{n} f_i x_i\right)^2\right]$ (Indirect method) or

  $\sigma^2 = \frac{h^2}{N^2}\left[N\sum_{i=1}^{n} f_i y_i^2 - \left(\sum_{i=1}^{n} f_i y_i\right)^2\right]$ (In shortcut method), where

  x_i = class marks of the class intervals, $\bar{x}$ = mean, $N = \sum_{i=1}^{n} f_i$, h = width of the class intervals, $y_i = \frac{x_i - A}{h}$, where A is the assumed mean.

- Standard deviation is the square root of variance and it is denoted by *s*. This means:

  $\text{Standard Deviation} = \sqrt{\text{Variance}}$

  Example:
  Find the variance and standard deviation for the following data.

Class	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	5	4	3	1	7

Solution:
Let assumed mean, A = 35
Here, h = 10, N = 20
We obtain the following table from the given data.

Class	**Frequency (f_i)**	**Mid-points (x_i)**	$y_i = \frac{x_i - 35}{10}$	y_i^2	$f_i y_i$	$f_i y_i^2$
10 – 20	5	15	–2	4	–10	20
20 – 30	4	25	–1	1	–4	4
30 – 40	3	35	0	0	0	0
40 – 50	1	45	1	1	1	1
50 – 60	7	55	2	4	14	28
Total	**N = 20**					

- The measure of variability, which is independent of units, is called the coefficient of variation. The coefficient of variation (C.V.) is defined as

  $$C.V = \frac{\sigma}{\bar{x}} \times 100, \bar{x} \neq 0$$

  Where, σ and $\bar{x}$ are standard deviation and mean of the data respectively.

- For comparing the variability or dispersion of two series, we first calculate the C.Vs of each series. The series having higher C.V. is said to be more variable than the other and the series having lower C.V. is said to be more consistent than the other.

- For two series with equal mean values, the series with greater standard deviation (or variance) is more variable or dispersed than the other. Also, the series with lower value of standard deviation (or variance) is said to be more consistent or less scattered than the other.

  Example:
  Which series, I or II, is more consistent?

	Series I	**Series II**
Mean	3100	3100
Variance	121	169

Solution:
Standard deviation of series I, $\sigma_1 = \sqrt{121} = 11$
Standard deviation of series II, $\sigma_2 = \sqrt{169} = 13$
Since the mean of both the series is the same, the series with lower standard deviation will be more consistent.
Thus, series I will be more consistent.