Chapter 15

STATISTICS

15.1 Overview

In earlier classes, you have studied measures of central tendency such as mean, mode, median of ungrouped and grouped data. In addition to these measures, we often need to calculate a second type of measure called a **measure of dispersion** which measures the **variation** in the observations about the **middle value**– mean or median etc.

This chapter is concerned with some important measures of dispersion such as mean deviation, variance, standard deviation etc., and finally analysis of frequency distributions.

15.1.1 *Measures of dispersion*

(a) **Range**The measure of dispersion which is easiest to understand and easiest to calculate is the **range**. Range is defined as:

Range = Largest observation – Smallest observation

(b) **Mean Deviation**

(i) Mean deviation for ungrouped data:

For n observation x_1, x_2, ..., x_n, the **mean deviation about their mean** $\bar{x}$ is given by

$$\text{M.D}\,(\bar{x}) = \frac{\sum |x_i - \bar{x}|}{n} \qquad (1)$$

Mean deviation about their median M is given by

$$\text{M.D (M)} = \frac{\sum |x_i - \text{M}|}{n} \qquad (2)$$

(ii) Mean deviation for discrete frequency distribution

Let the given data consist of discrete observations x_1, x_2, ... , x_n occurring with frequencies f_1, f_2, ... , f_n, respectively. In this case

$$\text{M.D}(\bar{x}) = \frac{\sum f_i |x_i - \bar{x}|}{\sum f_i} = \frac{\sum f_i |x_i - \bar{x}|}{\text{N}} \qquad (3)$$

$$\text{M.D (M)} = \frac{\sum f_i |x_i - \text{M}|}{\text{N}} \qquad (4)$$

where N = $\sum f_i$.

(iii) Mean deviation for continuous frequency distribution (Grouped data).

$$\text{M.D}(\bar{x}) = \frac{\sum f_i |x_i - \bar{x}|}{\text{N}} \qquad (5)$$

$$\text{M.D (M)} = \frac{\sum f_i |x_i - \text{M}|}{\text{N}} \qquad (6)$$

where x_i are the midpoints of the classes, $\bar{x}$ and M are, respectively, the mean and median of the distribution.

(c) **Variance :** Let $x_1, x_2, ..., x_n$ be n observations with $\bar{x}$ as the mean. The variance, denoted by σ^2, is given by

$$\sigma^2 = \frac{1}{n}\sum (x_i - \bar{x})^2 \qquad (7)$$

(d) **Standard Deviation:** If σ^2 is the variance, then σ, is called the standard deviation, is given by

$$\sigma = \sqrt{\frac{1}{n}\sum (x_i - \bar{x})^2} \qquad (8)$$

(e) **Standard deviation for a discrete frequency distribution** is given by

$$\sigma = \sqrt{\frac{1}{\text{N}}\sum f_i (x_i - \bar{x})^2} \qquad (9)$$

where f_i's are the frequencies of x_i' s and N = $\sum_{i=1}^{n} f_i$.

(f) **Standard deviation of a continuous frequency distribution (grouped data)** is given by

$$\sigma = \sqrt{\frac{1}{N}\sum f_i (x_i - \overline{x})^2} \qquad (10)$$

where x_i are the midpoints of the classes and f_i their respective frequencies. Formula (10) is same as

$$\sigma = \frac{1}{N}\sqrt{N\sum f_i x_i^2 - \left(\sum f_i x_i\right)^2} \qquad (11)$$

(g) Another formula for standard deviation :

$$\sigma_x = \frac{h}{N}\sqrt{N\sum f_i y_i^2 - \left(\sum f_i y_i\right)^2} \qquad (12)$$

where h is the width of class intervals and $y_i = \dfrac{x_i - A}{h}$ and A is the assumed mean.

15.1.2 ***Coefficient of variation*** It is sometimes useful to describe **variability** by expressing the standard deviation as a proportion of mean, usually a percentage. The formula for it as a percentage is

$$\text{Coefficient of variation} = \frac{\text{Standard deviation}}{\text{Mean}} \times 100$$

15.2 Solved Examples

Short Answer Type

Example 1 Find the mean deviation about the mean of the following data:

Size (x):	1	3	5	7	9	11	13	15
Frequency (f):	3	3	4	14	7	4	3	4

Solution Mean $= \overline{x} = \dfrac{\sum f_i x_i}{\sum f_i} = \dfrac{3+9+20+98+63+44+39+60}{42} = \dfrac{336}{42} = 8$

$$\text{M.D. } (\overline{x}) = \frac{\sum f_i |x_i - \overline{x}|}{\sum f_i} = \frac{3(7)+3(5)+4(3)+14(1)+7(1)+4(3)+3(5)+4(7)}{42}$$

$$= \frac{21 + 15 + 12 + 14 + 7 + 12 + 15 + 28}{42} = \frac{62}{21} = 2.95$$

Example 2 Find the variance and standard deviation for the following data:
57, 64, 43, 67, 49, 59, 44, 47, 61, 59

Solution Mean ($\bar{x}$) $= \dfrac{57 + 64 + 43 + 67 + 49 + 59 + 61 + 59 + 44 + 47}{10} = \dfrac{550}{10} = 55$

$$\text{Variance } (\sigma^2) = \frac{\sum (x_i - \bar{x})^2}{n}$$

$$= \frac{2^2 + 9^2 + 12^2 + 12^2 + 6^2 + 4^2 + 6^2 + 4^2 + 11^2 + 8^2}{10}$$

$$= \frac{662}{10} = 66.2$$

Standard deviation (σ) $= \sqrt{\sigma^2} = \sqrt{66.2} = 8.13$

Example 3 Show that the two formulae for the standard deviation of ungrouped data.

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \qquad \text{and} \qquad \sigma' = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$$

are equivalent.

Solution We have

$$\sum (x_i - \bar{x})^2 = \sum (x_i^2 - 2\bar{x}\, x_i + \bar{x}^2)$$

$$= \sum x_i^2 + \sum -2\,\bar{x}\, x_i + \sum \bar{x}^2$$

$$= \sum x_i^2 - 2\bar{x} \sum x_i + (\bar{x})^2 \sum 1$$

$$= \sum x_i^2 - 2\bar{x}\,(n\bar{x}) + n\,\bar{x}^2$$

$$= \sum x_i^2 - n\bar{x}^2$$

Dividing both sides by n and taking their square root, we get $\sigma = \sigma'$.

Example 4 Calculate **variance** of the following data :

Class interval	Frequency
4 - 8	3
8 - 12	6
12 - 16	4
16 - 20	7

$$\text{Mean } (\bar{x}) = \frac{\sum f_i x_i}{\sum f_i} = \frac{3\times 6 + 6\times 10 + 4\times 14 + 7\times 18}{20} = 13$$

Solution Variance $(\sigma^2) = \dfrac{\sum f_i (x_i - \bar{x})^2}{\sum f_i} = \dfrac{3(-7)^2 + 6(-3)^2 + 4(1)^2 + 7(5)^2}{20}$

$$= \frac{147 + 54 + 4 + 175}{20} = 19$$

Long Answer Type

Example 5 Calculate mean, variation and standard deviation of the following frequency distribution:

Classes	Frequency
1 - 10	11
10 - 20	29
20 - 30	18
30 - 40	4
40 - 50	5
50 - 60	3

Solution Let A, the assumed mean, be 25.5. Here $h = 10$

Classes	x_i	$y_i = \frac{x_i - 25.5}{10}$	f_i	$f_i y_i$	$f_i y_i^2$
1 - 10	5.5	–2	11	–22	44
10 - 20	15.5	–1	29	–29	29
20 - 30	25.5	0	18	0	0
30 - 40	35.5	1	4	4	4
40 - 50	45.5	2	5	10	20
50 - 60	55.5	3	3	9	27
			70	**–28**	**124**

$$x' = \frac{\sum f_i y_i}{\sum f_i} = \frac{-28}{70} = -0.4$$

Mean $= \bar{x} = 25.5 + (-10)(0.4) = 21.5$

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

$$= \frac{10 \times 10}{70 \times 70}\left[70(124) - (-28)^2\right]$$

$$= \frac{70\,(124)}{7 \times 7} - \frac{28 \times 28}{7 \times 7} = \frac{1240}{7} - 16 = 161$$

$$\text{S.D. } (\sigma) = \sqrt{161} = 12.7$$

Example 6 Life of bulbs produced by two factories A and B are given below:

Length of life (in hours)	Factory A (Number of bulbs)	Factory B (Number of bulbs)
550 - 650	10	8
650 - 750	22	60
750 - 850	52	24
850 - 950	20	16
950 - 1050	16	12
	120	**120**

The bulbs of which factory are more consistent from the point of view of length of life?

Solution Here h = 100, let A (assumed mean) = 800.

Length of life (in hour)	Mid values(x_i)	$y_i = \dfrac{x_i - A}{10}$	Factory A f_i	Factory A $f_i y_i$	Factory A $f_i y_i^2$	Factory B f_i	Factory B $f_i y_i$	Factory B $f_i y_i^2$
550 - 650	600	–2	10	–20	40	8	–16	32
650 - 750	700	–1	22	–22	22	60	– 60	60
750 - 850	800	0	52	0	0	24	0	0
850 - 950	900	1	20	20	20	16	16	16
950 - 1050	1000	2	16	32	64	12	24	48
			120	**10**	**146**	**120**	**–36**	**156**

For factory A

$$\text{Mean } (\bar{x}) = 800 + \frac{10}{120} \times 100 = 816.67 \text{ hours}$$

$$\text{S.D.} = \frac{100}{120}\sqrt{120\,(146) - 100} = 109.98$$

Therefore, Coefficient of variation (C.V.) = $\dfrac{\text{S.D.}}{\bar{x}} \times 100 = \dfrac{109.98}{816.67} \times 100 = 13.47$

For factory B

$$\text{Mean} = 800 + \left(\frac{-36}{120}\right) 100 = 770$$

$$\text{S.D.} = \frac{100}{120}\sqrt{120\,(156) - (-36)^2} = 110$$

Therefore, Coefficient of variation $= \dfrac{\text{S.D.}}{\text{Mean}} \times 100 = \dfrac{110}{770} \times 100 = 14.29$

Since C.V. of factory B > C.V. of factory A $\Rightarrow$ Factory B has more variability which means bulbs of factory A are more consistent.

Objective Type Questions

Choose the correct answer out of the four options given against each of the Examples 7 to 9 (M.C.Q.).

Example 7 The mean deviation of the data 2, 9, 9, 3, 6, 9, 4 from the mean is

(A) 2.23 (B) 2.57 (C) 3.23 (D) 3.57

Solution (B) is the correct answer

$$\text{M.D.}\,(\bar{x}) = \frac{\sum |x_i - \bar{x}|}{n} = \frac{4+3+3+3+0+3+2}{7} = 2.57$$

Example 8 Variance of the data 2, 4, 5, 6, 8, 17 is 23.33. Then variance of 4, 8, 10, 12, 16, 34 will be

(A) 23.23 (B) 25.33 (C) 46.66 (D) 48.66

Solution (C) is the correct answer. When each observation is multiplied by 2, then variance is also multiplied by 2.

Example 9 A set of n values $x_1, x_2, ..., x_n$ has standard deviation 6. The standard deviation of n values $x_1 + k, x_2 + k, ..., x_n + k$ will be

(A) σ (B) $\sigma + k$ (C) $\sigma - k$ (D) $k\sigma$

Solution (A) is correct answer. If each observation is increased by a constant k, then standard deviation is unchanged.

15.3 EXERCISE

Short Answer Type

1. Find the mean deviation about the mean of the distribution:

Size	20	21	22	23	24
Frequency	6	4	5	1	4

2. Find the mean deviation about the median of the following distribution:

Marks obtained	10	11	12	14	15
No. of students	2	3	8	3	4

3. Calculate the mean deviation about the mean of the set of first n natural numbers when n is an odd number.

4. Calculate the mean deviation about the mean of the set of first n natural numbers when n is an even number.

5. Find the standard deviation of the first n natural numbers.

6. The mean and standard deviation of some data for the time taken to complete a test are calculated with the following results:

Number of observations = 25, mean = 18.2 seconds, standard deviation = 3.25 seconds.

Further, another set of 15 observations $x_1, x_2, ..., x_{15}$, also in seconds, is now available and we have $\sum_{i=1}^{15} x_i = 279$ and $\sum_{i=1}^{15} x_i^2 = 5524$. Calculate the standard derivation based on all 40 observations.

7. The mean and standard deviation of a set of n_1 observations are $\bar{x}_1$ and s_1, respectively while the mean and standard deviation of another set of n_2 observations are $\bar{x}_2$ and s_2, respectively. Show that the standard deviation of the combined set of $(n_1 + n_2)$ observations is given by

$$\text{S.D.} = \sqrt{\frac{n_1(s_1)^2 + n_2(s_2)^2}{n_1 + n_2} + \frac{n_1 n_2 (\bar{x}_1 - \bar{x}_2)^2}{(n_1 + n_2)^2}}$$

8. Two sets each of 20 observations, have the same standard derivation 5. The first set has a mean 17 and the second a mean 22. Determine the standard deviation of the set obtained by combining the given two sets.

9. The frequency distribution:

x	A	2A	3A	4A	5A	6A
f	2	1	1	1	1	1

where A is a positive integer, has a variance of 160. Determine the value of A.

10. For the frequency distribution:

x	2	3	4	5	6	7
f	4	9	16	14	11	6

Find the standard distribution.

11. There are 60 students in a class. The following is the frequency distribution of the marks obtained by the students in a test:

Marks	0	1	2	3	4	5
Frequency	$x - 2$	x	x^2	$(x + 1)^2$	$2x$	$x + 1$

where x is a positive integer. Determine the mean and standard deviation of the marks.

12. The mean life of a sample of 60 bulbs was 650 hours and the standard deviation was 8 hours. A second sample of 80 bulbs has a mean life of 660 hours and standard deviation 7 hours. Find the overall standard deviation.

13. Mean and standard deviation of 100 items are 50 and 4, respectively. Find the sum of all the item and the sum of the squares of the items.

14. If for a distribution $\sum (x-5)=3,\ \sum (x-5)^2=43$ and the total number of item is 18, find the mean and standard deviation.

15. Find the mean and variance of the frequency distribution given below:

x	$1 \leq x < 3$	$3 \leq x < 5$	$5 \leq x < 7$	$7 \leq x < 10$
f	6	4	5	1

Long Answer Type

16. Calculate the mean deviation about the mean for the following frequency distribution:

Class interval	0 - 4	4 - 8	8 - 12	12 - 16	16 - 20
Frequency	4	6	8	5	2

17. Calculate the mean deviation from the median of the following data:

Class interval	0 - 6	6 - 12	12 - 18	18 - 24	24 - 30
Frequency	4	5	3	6	2

18. Determine the mean and standard deviation for the following distribution:

Marks	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Frequency	1	6	6	8	8	2	2	3	0	2	1	0	0	0	1

19. The weights of coffee in 70 jars is shown in the following table:

Weight (in grams)	**Frequency**
200 - 201	13
201 - 202	27
202 - 203	18
203 - 204	10
204 - 205	1
205 - 206	1

Determine variance and standard deviation of the above distribution.

20. Determine mean and standard deviation of first n terms of an A.P. whose first term is a and common difference is d.

21. Following are the marks obtained, out of 100, by two students Ravi and Hashina in 10 tests.

Ravi	25	50	45	30	70	42	36	48	35	60
Hashina	10	70	50	20	95	55	42	60	48	80

Who is more intelligent and who is more consistent?

22. Mean and standard deviation of 100 observations were found to be 40 and 10, respectively. If at the time of calculation two observations were wrongly taken as 30 and 70 in place of 3 and 27 respectively, find the correct standard deviation.

23. While calculating the mean and variance of 10 readings, a student wrongly used the reading 52 for the correct reading 25. He obtained the mean and variance as 45 and 16 respectively. Find the correct mean and the variance.

Objective Type Questions

Choose the correct answer out of the given four options in each of the Exercises 24 to 39 (M.C.Q.).

24. The mean deviation of the data 3, 10, 10, 4, 7, 10, 5 from the mean is

(A) 2 (B) 2.57 (C) 3 (D) 3.75

25. Mean deviation for n observations $x_1, x_2, ..., x_n$ from their mean $\overline{x}$ is given by

(A) $\sum_{i=1}^{n}(x_i - \overline{x})$ (B) $\frac{1}{n}\sum_{i=1}^{n}\left|x_i - \overline{x}\right|$

(C) $\sum_{i=1}^{n}(x_i - \overline{x})^2$ (D) $\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$

26. When tested, the lives (in hours) of 5 bulbs were noted as follows:
1357, 1090, 1666, 1494, 1623

The mean deviations (in hours) from their mean is

(A) 178 (B) 179 (C) 220 (D) 356

27. Following are the marks obtained by 9 students in a mathematics test:
50, 69, 20, 33, 53, 39, 40, 65, 59

The mean deviation from the median is:

(A) 9 (B) 10.5 (C) 12.67 (D) 14.76

28. The standard deviation of the data 6, 5, 9, 13, 12, 8, 10 is

(A) $\sqrt{\frac{52}{7}}$ (B) $\frac{52}{7}$ (C) $\sqrt{6}$ (D) 6

29. Let $x_1, x_2, ..., x_n$ be n observations and $\overline{x}$ be their arithmetic mean. The formula for the standard deviation is given by

(A) $\sum (x_i - \overline{x})^2$ (B) $\frac{\sum (x_i - \overline{x})^2}{n}$

(C) $\sqrt{\frac{\sum (x_i - \overline{x})^2}{n}}$ (D) $\sqrt{\frac{\sum x_i^2}{n} + \overline{x}^2}$

30. The mean of 100 observations is 50 and their standard deviation is 5. The sum of all squares of all the observations is

(A) 50000 (B) 250000 (C) 252500 (D) 255000

31. Let a, b, c, d, e be the observations with mean m and standard deviation s. The standard deviation of the observations $a + k$, $b + k$, $c + k$, $d + k$, $e + k$ is

(A) s (B) $k s$ (C) $s + k$ (D) $\frac{s}{k}$

32. Let x_1, x_2, x_3, x_4, x_5 be the observations with mean m and standard deviation s. The standard deviation of the observations $kx_1, kx_2, kx_3, kx_4, kx_5$ is

(A) $k + s$ (B) $\frac{s}{k}$ (C) $k s$ (D) s

33. Let $x_1, x_2, ... x_n$ be n observations. Let $w_i = lx_i + k$ for $i = 1, 2, ...n$, where l and k are constants. If the mean of x_i's is 48 and their standard deviation is 12, the mean of w_i's is 55 and standard deviation of w_i's is 15, the values of l and k should be

(A) $l = 1.25, k = -5$ (B) $l = -1.25, k = 5$

(C) $l = 2.5, k = -5$ (D) $l = 2.5, k = 5$

34. Standard deviations for first 10 natural numbers is

(A) 5.5 (B) 3.87 (C) 2.97 (D) 2.87

35. Consider the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. If 1 is added to each number, the variance of the numbers so obtained is

(A) 6.5 (B) 2.87 (C) 3.87 (D) 8.25

36. Consider the first 10 positive integers. If we multiply each number by –1 and then add 1 to each number, the variance of the numbers so obtained is

(A) 8.25 (B) 6.5 (C) 3.87 (D) 2.87

37. The following information relates to a sample of size 60: $\sum x^2 = 18000$,

$\sum x = 960$

The variance is

(A) 6.63 (B) 16 (C) 22 (D) 44

38. Coefficient of variation of two distributions are 50 and 60, and their arithmetic means are 30 and 25 respectively. Difference of their standard deviation is

(A) 0 (B) 1 (C) 1.5 (D) 2.5

39. The standard deviation of some temperature data in °C is 5. If the data were converted into ºF, the variance would be

(A) 81 (B) 57 (C) 36 (D) 25

Fill in the blanks in Exercises from 40 to 46.

40. Coefficient of variation = $\dfrac{\cdots}{\text{Mean}} \times 100$

41. If $\bar{x}$ is the mean of n values of x, then $\sum_{i=1}^{n}(x_i - \bar{x})$ is always equal to ________.

If a has any value other than $\bar{x}$, then $\sum_{i=1}^{n}(x_i - \bar{x})^2$ is __________ than $\sum (x_i - a)^2$

42. If the variance of a data is 121, then the standard deviation of the data is ________.

43. The standard deviation of a data is ____________ of any change in orgin, but is ______ on the change of scale.

44. The sum of the squares of the deviations of the values of the variable is ________ when taken about their arithmetic mean.

45. The mean deviation of the data is ________ when measured from the median.

46. The standard deviation is ________ to the mean deviation taken from the arithmetic mean.