S. K. Rawat

MEASURES OF DISPERSION

STATISTICS

12

MARKS

OBJECTIVES

- To revise and understand the meaning of data, class mark, class interval.
- To find the cumulative frequency of less than and more than type.
- To revise the concept of mean, median for raw, ungrouped and ungrouped data.
- To define and understand range, coefficient of range.
- To find range and coefficient of range for raw, ungrouped and grouped data.
- To find mean deviation from mean and median for raw, grouped and ungrouped data.
- To find standard deviation for raw, grouped and ungrouped data.
- To find standard deviation using step deviation method.
- To find variance and coefficient of variance.
- To compare two sets of observations using variance.

STATISTICAL TERMS

- **Data:** It is a collection of observations expressed in numerical figures obtained by measuring or counting.
- **Raw Data:** Data recorded in an arbitrary manner after their collection are called Raw data.
- **Tally Mark:** A upward slanted stroke (/) which is put against a value when it occurs once in raw data.
- **Frequency:** The total number of tally marks against each value is its frequency.
- **Ungrouped Data:** A data with frequency of its values in discrete form are called ungrouped data.

STATISTICAL TERMS

- **Grouped Data:** A tabular arrangement of raw data by putting the whole range of observations into a number of smaller groups or classes, showing respective class-frequencies against the class-intervals are called grouped data.

- **Class Limits:** The two end-values of a class-interval are called class limits.

- **Class Mark or Mid-value:** The class mark of the class-interval is the value exactly at the middle of the class-interval and is given by

$$\text{Class Mark} = \frac{\text{Lower class limit} + \text{Upper class limit}}{2}$$

RAW DATA

Marks of 50 students of a class in Mathematics

26	55	54	25	56
39	26	58	32	35
25	25	22	25	46
25	36	35	35	68
35	32	38	56	32
22	46	48	75	42
36	24	64	39	35
64	42	45	40	56
48	78	42	54	47
54	50	63	45	35

- Data presented in the table are **raw data**.
- Arranging the marks of data in ascending order.

22	32	36	46	56
22	32	38	46	56
24	32	39	47	56
25	35	39	48	58
25	35	40	48	63
25	35	42	50	64
25	35	42	54	64
25	35	42	54	68
26	35	45	54	75
26	36	45	55	78

UNGROUPED DATA

Marks of 50 students of a class in Mathematics

22	32	36	46	56
22	32	38	46	56
24	32	39	47	56
25	35	39	48	58
25	35	40	48	63
25	35	42	50	64
25	35	42	54	64
25	35	42	54	68
26	35	45	54	75
26	36	45	55	78

Marks	Tally Marks	Frequency
22	//	2
24	/	1
25	~~////~~	5
26	//	2
32	///	3
35	~~////~~ /	6
36	//	2
38	/	1
39	//	2
40	/	1
42	///	3
45	//	2
46	//	2
Total	--	32

Marks	Tally Marks	Frequency
47	/	1
48	//	2
50	/	1
54	///	3
55	/	1
56	///	3
58	/	1
63	/	1
64	//	2
68	/	1
75	/	1
78	/	1
Total	--	18

Total Frequency = 32 + 18 = 50

- Representing the data in simple frequency distribution table (***Ungrouped data***).
- Constructing a ***Grouped frequency distribution*** taking class-interval as 21-30, 31-40 and so on, the last class-interval being 71-80.

S. K. Rawat

GROUPED DATA

Marks	Tally Marks	Frequency	Marks	Tally Marks	Frequency
22	//	2	47	/	1
24	/	1	48	//	2
25	~~////~~	5	50	/	1
26	//	2	54	///	3
32	///	3	55	/	1
35	~~////~~ /	6	56	///	3
36	//	2	58	/	1
38	/	1	63	/	1
39	//	2	64	//	2
40	/	1	68	/	1
42	///	3	75	/	1
45	//	2	78	/	1
46	//	2	Total	--	18
Total	--	32			

Total Frequency = 32 + 18 = 50

- Constructing a ***Grouped frequency distribution*** taking class-interval as 21-30, 31-40 and so on, the last class-interval being 71-80.

Class-intervals	Tally Marks	Frequency (No. of Students)
21-30	~~////~~ ~~////~~	10
31-40	~~////~~ ~~////~~ ~~////~~	15
41-50	~~////~~ ~~////~~ /	11
51-60	~~////~~ ///	8
61-70	////	4
71-80	//	2
Total	--	50

DISCONTINUOUS GROUPED DATA

- The table shown is ***Grouped frequency distribution*** with **discontinuous** class-interval.
- The class-intervals can be made continuous by using correction factor.

Class-intervals	Tally Marks	Frequency (No. of Students)
21-30	~~////~~ ~~////~~	10
31-40	~~////~~ ~~////~~ ~~////~~	15
41-50	~~////~~ ~~////~~ /	11
51-60	~~////~~ ///	8
61-70	////	4
71-80	//	2
Total	--	50

$$\text{Correction Factor} = \frac{1}{2} \times d$$

where "d" is the difference between lower limit of upper class and upper limit of lower class.

CONTINUOUS GROUPED DATA

Discontinuous Class-interval

Class-intervals	Tally Marks	Frequency (No. of Students)
21-30		10
31-40		15
41-50		11
51-60		8
61-70		4
71-80		2
Total	--	50

Continuous Class-interval

Class-intervals	Tally Marks	Frequency (No. of Students)
20.5 - 30.5		10
30.5 – 40.5		15
40.5 – 50.5		11
50.5 – 60.5		8
60.5 – 70.5		4
70.5 – 80.5		2
Total	--	50

CUMULATIVE FREQUENCY

- In classification of statistical data, the number of values less than or more than a given value can be determined by accumulated frequency called as ***cumulative frequency***.
- Cumulative frequencies are of two types:
  - Less than cumulative frequencies.
  - More than cumulative frequencies.

CUMULATIVE FREQUENCY

To construct cumulative frequency distribution from Height of 80

Height (in cms)	110-119	120-129	130-139	140-149	150-159	Total
No. of students	7	15	30	20	8	80

- Here d = gap between any two consecutive classes = 1 and $\frac{1}{2}d = 0.5$
- Therefore the lower class boundary points are 100 – 0.5, 120 – 0.5,….etc and the last upper class boundary points is 159 + 0.5, i.e. the class boundary points are 109.5, 119.5,………., 159.5.

Class Intervals (Heights in cms)	Cumulative frequency	
	Less than	More than
< 109.5	0	80
109.5 - 119.5	7	80 – 7 = 73
119.5 - 129.5	7 + 15 = 22	73 – 15 = 58
129.5 - 139.5	22 + 30 = 52	58 – 30 = 28
139.5 - 149.5	52 + 20 = 72	28 – 20 = 8
149.5 - 159.5	72 + 8 = 80	8 – 8 = 0

REVISION: CENTRAL TENDENCY

- For a particular set of data, where a central value (average) is used to measure the entire data is called as measure of ***central tendency***.
- There are three measures of central tendency:
  - Mean
  - Median
  - Mode
- There are three types of mean:
  - Arithmetic Mean (A.M)
  - Geometric Mean (G.M)
  - Harmonic Mean (H.M)
- Of these, Arithmetic Mean is most commonly used. Infact for Mean we shall always refer to Arithmetic mean (A.M)

REVISION: MEAN

RAW DATA

- Arithmetic Mean $\left(\overline{x}\right)$

 The Arithmetic Mean for a given series of values, $x_1, x_2, \ldots, x_n$ is defined as the sum of these values divided by the total number of observations.

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \ldots\ldots + x_n}{N}$$

- **Example:** *Find the Arithmetic Mean of 7, 16, 24 and 53.*
- *Solution:* The required A.M

$$\overline{\mathrm{x}} = \frac{7+16+24+53}{4} = \frac{100}{4} = 25$$

REVISION: MEAN

UNGROUPED DATA

- Mean $\left(\bar{x}\right)$

If $x_1, x_2, \ldots, x_n$ are the values of variable x and $f_1, f_2, \ldots, f_n$ are their respective frequencies, then mean is defined as

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \ldots\ldots + f_nx_n}{f_1 + f_2 + f_3 + \ldots\ldots + f_n} = \frac{\sum f_i x_i}{\sum f_i} = \frac{\sum f_i x_i}{N}$$

REVISION: MEAN

GROUPED DATA

- **Example:** *Find the Mean for following distribution.*

Marks	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50
No. of students	5	10	25	20	4

- *Solution:* The required Mean is calculated as $\bar{x} = \dfrac{\sum f_i x_i}{N}$

Marks	Mid-marks (x_i)		$F_i x_i$
0 – 10	5	5	25
10 – 20	15	10	150
20 – 30	25	25	625
30 – 40	35	20	700
40 – 50	45	4	180
		$N = \sum f = 64$	$\sum fx = 1680$

$$\bar{x} = \frac{\sum f_i x_i}{N}$$

$$\therefore \quad \bar{x} = \frac{1680}{64}$$

$$\therefore \quad \bar{x} = 26.25$$

REVISION: MEDIAN

RAW DATA

- Median (M)
- It is the value of the variable that divides the series into two equal parts such that half of the total number of values are less than or equal to it and half are greater than or equal to it.
- The given values are arranged in ascending order.
- If ***x*** is variable with ***'N'*** number of observations, then ***'N'*** can be even or order.
- If ***N = even number***, then

$$\text{Median} = \frac{\left(\frac{N}{2}\right)^{th}\text{place observation} + \left(\frac{N}{2}+1\right)^{th}\text{place observation}}{2}$$

MEDIAN FOR RAW DATA

Example: The marks obtained by 12 students out of 50 marks are:

25, 24, 23, 32, 40, 27, 30, 25, 20, 10, 15, 45.

Solution:

Arranging the given values (i.e., marks) in ascending order, we get

10, 15, 20, 23, 24, 25, 25, 27, 30, 32, 40, 45.

Here N = number of values = 12 which is even.

$\therefore$ There are ***two middle terms*** which are the 6^{th} and 7^{th} terms.

$$\text{Median} = \frac{6^{th}\text{ term} + 7^{th}\text{ term}}{2} = \frac{25+25}{2} = 25$$

MEDIAN FOR RAW DATA

- If ***N = odd number***, then

$$\text{Median} = \left(\frac{N+1}{2}\right)^{th} \text{ place observation}$$

Example: Find the median of 100, 90, 40, 50, 125, 65, 110.

Solution: Arranging the given values in ascending order,

i.e. 40, 50, 65, **90**, 100, 110, 125.

There are 7 values, i.e., ***N =7*** which is odd.

$\therefore$ There is only ***one middle*** value which is the $\left(\frac{N+1}{2}\right)^{th}$ term.

i.e., $\left(\frac{7+1}{2}\right)^{th} = 4^{th}$ term. Also 4^{th} value is 90.

Median = 90.

MEDIAN FOR UNGROUPED DATA

- In this case, first calculate ***cumulative frequency (“less than type”)*** corresponding to each value of variable.

- Then the median is the value of variable corresponding to the cumulative frequency depending on the value of N, where $N = \sum f$ = total frequency.

- If ***N is odd***, then the median is the value of variable corresponding to the cumulative frequency $\frac{(N+1)}{2}$

- But if ***N is even***, then the median is arithmetic mean of two middle values corresponding to cumulative frequencies

$$\text{i.e. } \frac{\frac{N}{2} + \left(\frac{N}{2} + 1\right)}{2}$$

MEDIAN FOR UNGROUPED DATA

Example: Find the median for following data.

Weight in lbs	115	120	130	140	145
No. of men	8	20	35	24	12

Solution: Calculation of cumulative frequency

x_i	Frequency f_i	Cumulative frequency
115	8	8
120	20	20 + 8 = 28
130	35	35 + 28 = 63
140	24	24 + 63 = 87
145	12	12 + 87 = 99 = N
Total	N = 99	

Here N = 99

which is ***odd***

$$\therefore \frac{(N+1)}{2} = \frac{99+1}{2} = 50$$

As 50 lies between 28 and 63.

Hence Median = the value corresponding to cumulative frequency 50 = **130 lbs.**

MEDIAN FOR UNGROUPED DATA

Example: Find the median for following data.

Class Test Marks	18	20	21	22	23	24
No. of students	10	23	32	28	12	5

Solution: Calculation of cumulative frequency

Here N = 110 which is ***even***

x_i	Frequency f_i	Cumulative frequency
18	10	10
20	23	23 + 10 = 33
21	32	32 + 33 = 65
22	28	28 + 65 = 93
23	12	12 + 93 = 105
24	5	5 + 105 = 110 = N
Total	N = 110	

$$\therefore \left(\frac{N}{2}\right)^{th} = \left(\frac{110}{2}\right)^{th} = 55^{th}$$

$$\& \left(\frac{N}{2}+1\right)^{th} = \left(\frac{110}{2}+1\right)^{th} = 56^{th}$$

As 55^{th} & 56^{th} lies between 33 and 65.

$$\therefore \text{Median} = \frac{55^{th}\ \text{value} + 56^{th}\ \text{value}}{2} = \frac{21+21}{2} = \frac{42}{2} = 21$$

Hence Median = the value corresponding to cumulative frequency 65 = **21 marks.**

MEDIAN FOR GROUPED DATA

- In this case, first calculate ***cumulative frequency*** ***("less than type")*** corresponding to each value of variable.

- Then Median is the value of the variable which corresponds to the cumulative frequency N/2 and the class in which the median lies is called the median class

- The median is calculated as $$Median = l_1 + \frac{\frac{N}{2} - f_c}{f_m} \times c$$

  - *where* l_1 = *lower boundary of median class.*
  - $N = \sum f$ = *total frequency.*
  - f_c = *cumulative frequency below* l_1 .
  - f_m = *frequency of the median class.*
  - c = *width of median class.*

MEDIAN FOR GROUPED DATA

Example: Find the median for following data.

Marks	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50
No. of students	8	20	36	24	12

Solution: Calculation of cumulative frequency

Class-interval	Frequency f	Cumulative frequency
0 - 10	8	8
10 -20	20	20 + 8 = 28 = f_c
l_1 = 20 - 30	36 = f_m	36 + 28 = 64
30 - 40	24	24 + 64 = 88
40 - 50	12	12 + 88 = 100 = N

Here N = 100

$$\therefore \frac{N}{2} = \frac{100}{2} = 50$$

From table *c = 10*

$$\therefore M = 20 + \frac{50-28}{36} \times 10$$

$$\therefore \text{Median} = 20 + 6.11 = 26.11$$

Median = the value corresponding to cumulative frequency 50.

∴ Median class is 20 – 30 and Median is calculated as $Median = l_1 + \dfrac{\frac{N}{2} - f_c}{f_m} \times c$

MEASURES OF DISPERSION

- The word dispersion means deviation or difference.
- In statistics, dispersion refers to deviation of the values of a variable from their central value.
- Measures of dispersion indicate the extent to which individual observations vary from their averages i.e. mean, median or mode.

Definition:

- "*Dispersion is the measure of deviation or variation of the variables about a central value*".

MEASURES OF DISPERSION

Absolute and relative measures of dispersion:

- The dispersion of a series may be measured either absolutely or relatively.
- If the dispersion is expressed in terms of the original units of the series, it is called absolute measure of dispersion.
- So for comparison point of view it is necessary to calculate the relative measures of dispersion which are expressed as percentage form. These types of expressions are called coefficients of dispersion.

MEASURES OF DISPERSION

Types of measures of dispersion:

- The following are the important measures of dispersion.
  - Range
  - Mean absolute deviation or Mean deviation
  - Standard deviation

RANGE

Range:

- It is the difference between the largest and the smallest value of observation of data.

For Ungrouped Data:

- The difference between the largest and the smallest values of the variable is called the range.
  - If L is the largest value and S is the smallest value of variable then, **Range = L – S**.

Example: Consider the data 12, 14, 11, 9, 15. Here, the largest value is 15 and the smallest value is 9.

Range = 15 – 9 = 6.

RANGE

For Grouped Data:

- The difference between the upper class limit of highest class and the lower class limit of lower class is called the range.
- The difference between the mid-value of highest class interval and the mid-value of lower class interval is called the range.
  - If L is the largest value and S is the smallest value of variable then, **Range = L – S**.

RANGE

Example: Consider the following distribution.

Marks	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59
No. of students	6	10	16	14	8

Solution: Making the class boundaries continuous, we get

Marks	9.5 – 19.5	19.5 – 29.5	29.5 – 39.5	39.5 – 49.5	49.5 – 59.5
No. of students	6	10	16	14	8

$\therefore$ S = Lower limit of lowest class = 9.5

$\therefore$ L = Upper limit of highest class = 59.5

Then, ***Range = L – S = 59.5 – 9.5 = 50***.

COEFFICIENT OF RANGE

Definition:

- It is defined as the ratio of Range to sum of highest and lowest class values.

$$\text{Coefficient of Range} = \frac{\text{Range}}{\text{Sum of highest \& lowest values}}$$

$$\text{Coefficient of Range} = \frac{L-S}{L+S}$$

COEFFICIENT OF RANGE

Example: Consider the following distribution.

Marks	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59
No. of students	6	10	16	14	8

Solution: Making the class boundaries continuous, we get

Marks	9.5 – 19.5	19.5 – 29.5	29.5 – 39.5	39.5 – 49.5	49.5 – 59.5
No. of students	6	10	16	14	8

$\therefore$ S = Lower limit of lowest class = 9.5

$\therefore$ L = Upper limit of highest class = 59.5

Then, *Range = L – S = 59.5 – 9.5 = 50*.

Coefficient of Range = $\frac{L-S}{L+S}=\frac{59.5-9.5}{59.5+9.5}=\frac{50}{69}=0.725$

MEAN DEVIATION

Mean Deviation:

- Mean deviation of a series of values of a variable is the arithmetic mean of all the absolute deviations from any one of its averages (mean, median or mode).
- It is a absolute measure of dispersion.

$$\text{Thus, Mean deviation} = \frac{\sum |d_i|}{N}$$

$$\text{where } |d_i| = |x_i - \bar{x}| \qquad \text{where } \bar{x} = \text{arithmetic mean}$$

$$\text{OR} \qquad \&$$

$$|d_i| = |x_i - M| \qquad M = \text{Median}$$

MEAN DEVIATION

RAW DATA

- **Mean Deviation about Mean** $= \dfrac{\sum |d_i|}{N} = \dfrac{\sum |x_i - \bar{x}|}{N}$
- **Mean Deviation about Median** $= \dfrac{\sum |d_i|}{N} = \dfrac{\sum |x_i - M|}{N}$

UNGROUPED FREQUENCY DISTRIBUTION OR DISCRETE DATA

- **Mean Deviation about Mean** $= \dfrac{\sum f_i |d_i|}{N} = \dfrac{\sum f_i |x_i - \bar{x}|}{\sum f_i}$
- **Mean Deviation about Median** $= \dfrac{\sum f_i |d_i|}{N} = \dfrac{\sum f_i |x_i - M|}{\sum f_i}$

MEAN DEVIATION

GROUPED FREQUENCY DISTRIBUTION

- **Mean Deviation about Mean** $= \dfrac{\sum f_i |d_i|}{N} = \dfrac{\sum f_i |x_i - \bar{x}|}{\sum f_i}$

$$\textit{where } x_i = \textit{mid value or class value}$$

- **Mean Deviation about Median** $= \dfrac{\sum f_i |d_i|}{N} = \dfrac{\sum f_i |x_i - M|}{\sum f_i}$

$$\textit{where } M = \textit{median of distribution}$$

Example

Example: Calculate the mean deviation about the mean.

Marks	3	4	5	6	7	8
No. of students	1	3	7	5	2	2

Solution: Table to obtain mean deviation about the mean

x_i	f_i	$f_i x_i$	$\lvert d_i \rvert = \lvert x_i - \ddot{x} \rvert$	$f_i \lvert d_i \rvert$
3	1	3	2.5	2.5
4	3	12	1.5	4.5
5	7	35	0.5	3.5
6	5	30	0.5	2.5
7	2	14	1.5	3
8	2	16	2.5	5
	$\sum f_i = 20$	$\sum f_i x_i = 110$		$\sum f_i d_i = 21$

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{110}{20} = 5.5$$

$$\text{Now, Mean deviation about mean} = \frac{\sum f_i \lvert d_i \rvert}{N} = \frac{21}{20} = 1.05$$

Example

Example: Calculate the mean deviation about the mean.

Marks	**0 - 10**	**10 - 20**	**20 - 30**	**30 - 40**	**40 - 50**
No. of students	**5**	**8**	**15**	**16**	**6**

Solution: Table to obtain mean deviation about the mean

Class Interval	**Mid-value (x_i)**	**f_i**	**$f_i x_i$**	**$\lvert d_i \rvert = \lvert x_i - \ddot{x} \rvert$**	**$f_i \lvert d_i \rvert$**
0 – 10	**5**	**5**	**25**	**22**	**110**
10 – 20	**15**	**8**	**120**	**12**	**96**
20 – 30	**25**	**15**	**375**	**2**	**30**
30 – 40	**35**	**16**	**560**	**8**	**128**
40 - 50	**45**	**6**	**270**	**18**	**108**
		$\sum f_i = 50$	**$\sum f_i x_i = 1350$**		**$\sum f_i \lvert d_i \rvert = 472$**

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

$$= \frac{1350}{50} = 27$$

$$\text{Now, Mean deviation about mean} = \frac{\sum f_i |d_i|}{N} = \frac{472}{50} = 9.44$$

Example

Solution: Table to find mean deviation about the mean using ***step deviation method***

Class Interval	Mid-value (x_i)	Frequency f_i	$u_i = \frac{x_i - a}{c}$	$f_i u_i$	$\lvert d_i \rvert = \lvert x_i - \ddot{x} \rvert$	$f_i \lvert d_i \rvert$
0 – 10	5	5	-2	-10	22	110
10 – 20	15	8	-1	-8	12	96
20 – 30	25	15	0	0	2	30
30 – 40	35	16	1	16	8	128
40 - 50	45	6	2	12	18	108
		$\sum f_i = 50$		$\sum f_i u_i = 10$		$\sum f_i \lvert d_i \rvert = 472$

where a = 25
and c = 10

$$\therefore \ \overline{U} = \frac{\sum f_i u_i}{\sum f_i} = \frac{10}{50}$$

$$\text{Mean, } \bar{x} = a + \left(\overline{U} \times c\right) = 25 + \left(\frac{10}{50} \times 10\right) = 25 + 2 = 27$$

$$\text{Now, Mean deviation about mean} = \frac{\sum f_i \lvert d_i \rvert}{\sum f_i} = \frac{472}{50} = 9.44$$

Example

Example: Calculate the mean deviation about the median.

Weights (gms)	10 - 15	15 - 20	20 - 25	25 - 30	30 - 35	35 - 40	40 - 45
No. of items	7	12	16	25	19	15	6

Solution: Table to obtain mean deviation about the median

Class Interval	Mid-value (x_i)	f_i	C.F. Less than
10 – 15	12.5	7	7
15 – 20	17.5	12	19
20 – 25	22.5	16	35 = f_c
l_1 25 – 30	27.5	25 = f_m	60
30 - 35	32.5	19	79
35 – 40	37.5	15	94
40 - 45	42.5	6	100
Total		$\sum f_i = 100$	

Here, $\sum f_i = N = 100$ $\quad \therefore \frac{N}{2} = \frac{100}{2} = 50$

$\therefore$ Median class $= 25-30$ and

$f_c = 35,\ l_1 = 25,\ f_m = 25,\ c = 5$

$$\therefore \text{Median} = l_1 + \frac{\frac{N}{2} - f_c}{f_m} \times c$$

$$= 25 + \frac{50-35}{25} \times 5$$

$$= 25 + 3$$

$$= 28$$

Example

Solution: Table to obtain mean deviation about the median

Class Interval	Mid-value (x_i)	f_i	C.F. Less than	$\|d_i\| = \|x_i - M\|$	$f_i\|d_i\|$
10 – 15	12.5	7	7	15.5	108.5
15 – 20	17.5	12	19	10.5	126.00
20 – 25	22.5	16	35	5.5	88.00
25 – 30	27.5	25	60	0.5	12.50
30 - 35	32.5	19	79	4.5	85.50
35 – 40	37.5	15	94	9.5	142.50
40 - 45	42.5	6	100	14.5	87.00
Total		$\sum f_i = 100$			$\sum f_i\|d_i\| = 650$

Here,

$\sum f_i = N = 100$

Thus, $M = 28$

Now, Mean deviation about median $= \dfrac{\sum f_i |d_i|}{N} = \dfrac{650}{100} = 6.5$

STANDARD DEVIATION

Standard Deviation:

- Standard deviation is defined as the square root of the mean of the squares of deviations from mean.
- S.D is denoted as sigma (σ).

Steps to find 'σ':

- Find mean of the observations directly or using step deviation methods.
- Calculate sum of squares of mean deviation for all values $i.e \sum \left(x_i - \bar{x}\right)^2$
- Find mean of $\frac{\sum \left(x_i - \bar{x}\right)^2}{N}$
- Take square root of this mean, i.e., **σ = S.D =** $\sqrt{\frac{\sum \left(x_i - \bar{x}\right)^2}{N}}$

STANDARD DEVIATION

RAW DATA

- **Standard Deviation = S.D. =**

$$\sigma = \sqrt{\frac{\sum \left(x_i - \overline{x}\right)^2}{N}} = \sqrt{\frac{\sum d_i^{\,2}}{N}}$$

$$where \quad d_i = x_i - \overline{x} \qquad \overline{x} = \frac{\sum x_i}{N} \qquad N = \sum f_i$$

UNGROUPED FREQUENCY DISTRIBUTION OR UNGROUPED DATA

- **Standard Deviation =**

$$\sigma = \sqrt{\frac{\sum f_i \left(x_i - \overline{x}\right)^2}{N}} = \sqrt{\frac{\sum f_i d_i^{\,2}}{N}} = \sqrt{\frac{\sum f_i x_i^{\,2}}{N} - \left(\overline{x}\right)^2}$$

$$where \quad d_i = x_i - \overline{x} \qquad \overline{x} = \frac{\sum f_i x_i}{N} \qquad N = \sum f_i$$

STANDARD DEVIATION

GROUPED FREQUENCY DISTRIBUTION

- **Standard Deviation = S.D**

$$\sigma = \sqrt{\frac{\sum f_i\left(x_i - \bar{x}\right)^2}{N}} = \sqrt{\frac{\sum f_i d_i^{\,2}}{N}} = \sqrt{\frac{\sum f_i x_i^{\,2}}{N} - \left(\bar{x}\right)^2}$$

$$\text{where } x_i = \text{mid} - \text{value or class} - \text{mark, } d_i = x_i - \bar{x}$$

$$\bar{x} = \frac{\sum f_i x_i}{N} \text{ and } N = \sum f_i$$

- The ratio of standard deviation to mean is called ***coefficient of standard deviation.***

$$\text{Coefficient of S.D.} = \frac{\sigma}{\text{Mean}} = \frac{\sigma}{\bar{x}}$$

VARIANCE

- The square of standard deviation is called ***variance.***

$$\sigma^2 = \frac{\sum \left(x_i - \overline{x}\right)^2}{N} = \frac{\sum d_i^{\,2}}{N} \;.................\; \textit{Raw data}$$

$$\sigma^2 = \frac{\sum f_i \left(x_i - \overline{x}\right)^2}{N} = \frac{\sum f_i d_i^{\,2}}{N} = \frac{\sum f_i x_i^{\,2}}{N} - \left(\overline{x}\right)^2 \;...........\; \textit{Frequency distribution}$$

- The ratio of standard deviation to mean in percentage is called ***coefficient of variance.***

$$\text{Coefficient of variance} = \frac{\text{S.D}}{\text{Mean}} \times 100 = \frac{\sigma}{\text{Mean}} \times 100 = \frac{\sigma}{\overline{\text{x}}} \times 100$$

Example

Example: Calculate the standard deviation and variance for the following data. 20, 70, 35, 60, 40, 25, 55, 15.

Solution: Arranging data in ascending order 15, 20, 25, 35, 40, 55, 60, 70.

Table to obtain S.D.

x_i	$d_i = x_i - \ddot{x}$	d_i^2
15	- 25	625
20	- 20	400
25	- 15	225
35	- 5	25
40	0	0
55	15	225
60	20	400
70	30	900
$\sum x_i = 320$		$\sum d_i^2 = 2800$

$N = 8$ $\quad \therefore \bar{x} = \dfrac{\sum x_i}{N} = \dfrac{320}{8} = 40$

$$\therefore \text{ S.D.} = \sigma = \sqrt{\frac{\sum d_i^2}{N}} = \sqrt{\frac{2800}{8}} = 18.708$$

$$\therefore \text{ Variance} = \sigma^2 = \frac{\sum d_i^2}{N} = \frac{2800}{8} = 350$$

Example

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

Weekly Expenses below Rs.	5	10	15	20	25
No. of students	6	16	28	38	46

Solution: Table to obtain standard deviation

x_i	f_i	$f_i x_i$	$d_i = x_i - \ddot{x}$	d_i^2	$f_i d_i^2$
5	6	30	- 13.8	190.44	1142.64
10	16	160	- 8.8	77.44	1230.04
15	28	420	- 3.8	14.44	404.32
20	38	760	1.2	1.44	54.72
25	46	1150	6.2	38.44	1768.24
	$\sum f_i = 134$	$\sum f_i x_i = 2520$			$\sum f_i d_i^2 = 4608.96$

$$N = \sum f_i = 134$$

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{2520}{134} = 18.8$$

$$\therefore \text{S.D.} = \sigma = \sqrt{\frac{\sum f_i d_i^2}{N}} = \sqrt{\frac{4608.96}{134}} = 5.8$$

$$\therefore \text{Variance} = \sigma^2 = \frac{\sum f_i d_i^2}{N} = \frac{4608.96}{134} = 34.4$$

Example

Example: Find the standard deviation for the following data.

C.I	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100
f_i	20	130	220	70	60

Solution: Table to obtain standard deviation

C.I	x_i	f_i	$f_i x_i$	$d_i = x_i - \ddot{x}$	d_i^2	$f_i d_i^2$
0 - 20	10	20	200	- 40.8	1664.64	33292.8
20 - 40	30	130	3900	- 20.8	432.64	56243.2
40 - 60	50	220	11000	- 0.8	0.64	140.8
60 - 80	70	70	4900	19.2	368.64	25804.8
80 - 100	90	60	5400	39.2	1536.64	92198.4
		$\sum f_i = N = 500$	$\sum f_i x_i = 25400$			$\sum f_i d_i^2 = 207679.4$

$$N = \sum f_i = 500$$

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{25400}{500} = 50.8$$

$$\therefore \text{ S.D.} = \sigma = \sqrt{\frac{\sum f_i d_i^2}{N}} = \sqrt{\frac{207679.4}{500}} = 20.38$$

Example

Example: Calculate the mean, S.D, variance coefficient of standard deviation and coefficient of variation for the following data.

Expenses Rs.	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
Frequency	14	23	27	21	15

Solution:

$$N = \sum f_i = 100$$

Mean,

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

$$= \frac{2500}{100} = 25$$

C.I	x_i	f_i	$f_i x_i$	$d_i = x_i - \ddot{x}$	d_i^2	$f_i d_i^2$
0 – 10	5	14	70	- 20	400	5600
10 – 20	15	23	345	- 10	100	2300
20 – 30	25	27	675	0	0	0
30 – 40	35	21	735	10	100	2100
40 – 50	45	15	675	20	400	6000
		$\sum f_i = N = 100$	$\sum f_i x_i = 2500$			$\sum f_i d_i^2 = 16000$

Example

Solution: contd………..

From the table $\sum f_i = N = 100$, **Mean (ẍ) = 25** and $\sum f_i d_i^2 = 16000$

$$\therefore \ \text{S.D.} = \sigma = \sqrt{\frac{\sum f_i d_i^2}{N}} = \sqrt{\frac{16000}{100}} = 12.65$$

$$\therefore \ \text{Variance} = \sigma^2 = \frac{\sum f_i d_i^2}{N} = \frac{16000}{100} = 160$$

$$Coefficient\ of\ S.D. = \frac{\sigma}{\overline{x}} = \frac{12.65}{25} = 0.506$$

$$\text{Coefficient of variation} = \frac{\sigma}{\overline{x}} \times 100 = 0.506 \times 100 = 50.6$$

COMPARISON OF TWO SETS OF OBSERVATIONS

- If there are two sets, set I and set II having respective means and standard deviation.

STEPS

- *Find coefficient of variance for each set.*
- *Now compare coefficient of variance.*
- *The coefficient of variance which has least is more reliable or consistent.*

Example

Example: In two factories A and B, engaged in the same area of the industry, the average weekly wages (in Rs.) and the S.D are as below.

Factory	Average wages	S.D
A	34.5	5.0
B	28.5	4.5

Which factory A or B has greater variability in individual wages?

Solution:

Factory	Average wages $\ddot{x}$	S.D σ	Coefficient of variance = $\frac{\sigma}{\overline{X}} \times 100$
A	34.5	5.0	$V_A = 14.50\ \%$
B	28.5	4.5	$V_B = 15.79\ \%$

As we observe $V_A < V_B$.
Therefore **factory A** is ***more consistent*** and ***less variable*** than **factory B**.

Example

Example: The two sets of observations are given below.

SET I	SET II
ẍ = 82.5	ẍ = 48.75
σ = 7.3	σ = 8.35

Which of two is more consistent?

Solution:

SET I	SET II
ẍ = 82.5	ẍ = 48.75
σ = 7.3	σ = 8.35
$V_A = 8.848\ \%$	$V_B = 17.12\ \%$

For Set I

$$V_A = \frac{\sigma}{\bar{x}} \times 100 = \frac{7.3}{82.5} \times 100$$

$$V_A = 8.848\%$$

For Set II

$$V_B = \frac{\sigma}{\bar{x}} \times 100 = \frac{8.35}{48.75} \times 100$$

$$V_B = 17.12\%$$

As we observe $V_A < V_B$.

Therefore **Set I** is ***more consistent*** and ***less variable*** than **Set II**.