CHAPTER 6

Measures of Dispersion



Studying this chapter should enable you to:

- *know the limitations of averages;*
- *appreciate the need for measures of dispersion;*
- *enumerate various measures of dispersion;*
- *calculate the measures and compare them;*
- *distinguish between absolute and relative measures.*

1. INTRODUCTION

In the previous chapter, you have studied how to sum up the data into a single representative value. However, that value does not reveal the variability present in the data. In this chapter you will study those measures, which seek to quantify variability of the data.

Three friends, Ram, Rahim and Maria are chatting over a cup of tea. During the course of their conversation, they start talking about their family incomes. Ram tells them that there are four members in his family and the average income per member is Rs 15,000. Rahim says that the average income is the same in his family, though the number of members is six. Maria says that there are five members in her family, out of which one is not working. She calculates that the average income in her family too, is Rs 15,000. They are a little surprised since they know that Maria's father is earning a huge salary. They go into details and gather the following data:

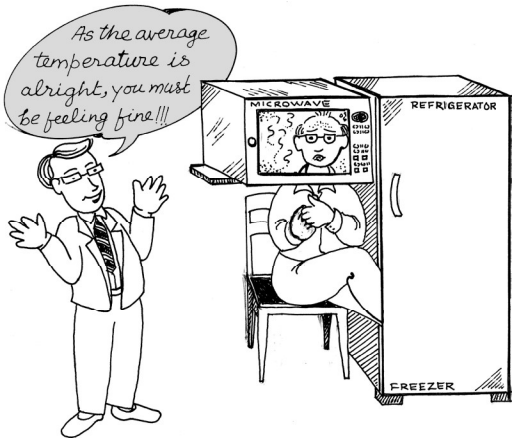
Family Incomes

Sl. No.	*Ram*	*Rahim*	*Maria*
1.	12,000	7,000	0
2.	14,000	10,000	7,000
3.	16,000	14,000	8,000
4.	18,000	17,000	10,000
5.	-----	20,000	50,000
6.	-----	22,000	------
Total income	*60,000*	*90,000*	*75,000*
Average income	*15,000*	*15,000*	*15,000*

Do you notice that although the average is the same, there are considerable differences in individual incomes?

It is quite obvious that averages try to tell only one aspect of a distribution i.e. a representative size of the values. To understand it better, you need to know the spread of values also.

You can see that in Ram's family, differences in incomes are comparatively lower. In Rahim's family, differences are higher and in Maria's family, the differences are the highest. Knowledge of only average is insufficient. If you have another value which reflects the quantum of variation in values, your understanding of a distribution improves considerably. For example, per capita income gives only the average income. A measure of dispersion can tell you about income inequalities, thereby improving the understanding of the relative standards of living enjoyed by different strata of society.



Dispersion is the extent to which values in a distribution differ from the average of the distribution.

To quantify the extent of the variation, there are certain measures namely:

(i) Range
(ii) Quartile Deviation
(iii) Mean Deviation
(iv) Standard Deviation

Apart from these measures which give a numerical value, there is a graphic method for estimating dispersion.

Range and Quartile Deviation measure the dispersion by calculating the spread within which the values lie. Mean Deviation and Standard Deviation calculate the extent to which the values differ from the average.

2. Measures Based upon Spread of Values

Range

Range (R) is the difference between the largest (L) and the smallest value (S) in a distribution. Thus,

$R = L - S$

Higher value of Range implies higher dispersion and vice-versa.

Activities

Look at the following values:
20, 30, 40, 50, 200

- Calculate the Range.
- What is the Range if the value 200 is not present in the data set?
- If 50 is replaced by 150, what will be the Range?

Range: Comments

Range is unduly affected by extreme values. It is not based on all the values. As long as the minimum and maximum values remain unaltered, any change in other values does not affect range. It cannot be calculated for open-ended frequency distribution.

Notwithstanding some limitations, Range is understood and used frequently because of its simplicity. For example, we see the maximum and minimum temperatures of different cities almost daily on our TV screens and form judgments about the temperature variations in them.

Open-ended distributions are those in which either the lower limit of the lowest class or the upper limit of the highest class or both are not specified.

Activity

- Collect data about 52-week high/ low of shares of 10 companies from a newspaper. Calculate the range of share prices. Which company's share is most volatile and which is the most stable?

Quartile Deviation

The presence of even one extremely high or low value in a distribution can reduce the utility of range as a measure of dispersion. Thus, you may need a measure which is not unduly affected by the outliers.

In such a situation, if the entire data is divided into four equal parts, each containing 25% of the values, we get the values of Quartiles and Median. (You have already read about these in Chapter 5).

The upper and lower quartiles (Q_3 and Q_1, respectively) are used to calculate Inter Quartile Range which is $Q_3 - Q_1$.

Inter-Quartile Range is based upon middle 50% of the values in a distribution and is, therefore, not affected by extreme values. Half of the Inter-Quartile Range is called Quartile Deviation (Q.D.). Thus:

$$Q.D. = \frac{Q_3 - Q_1}{2}$$

Q.D. is therefore also called *Semi-Inter Quartile Range.*

Calculation of Range and Q.D. for ungrouped data

Example 1

Calculate Range and Q.D. of the following observations:

20, 25, 29, 30, 35, 39, 41, 48, 51, 60 and 70

Range is clearly 70 – 20 = 50

For Q.D., we need to calculate values of Q_3 and Q_1.

© BSE&C NOT TO BE REPUBLISHED

Q_1 is the size of $\frac{n+1}{4}$th value.

n being 11, Q_1 is the size of 3rd value.

As the values are already arranged in ascending order, it can be seen that Q_1, the 3rd value is 29. [What will you do if these values are not in an order?]

Similarly, Q_3 is size of $\frac{3(n+1)}{4}$th value; i.e. 9th value which is 51. Hence $Q_3 = 51$

$$Q.D. = \frac{Q_3 - Q_1}{2} = \frac{51-29}{2} = 11$$

Do you notice that Q.D. is the average difference of the Quartiles from the median.

Activity

- Calculate the median and check whether the above statement is correct.

Calculation of Range and Q.D. for a frequency distribution.

Example 2

For the following distribution of marks scored by a class of 40 students, calculate the Range and Q.D.

TABLE 6.1

Class intervals C I	*No. of students* (f)
0–10	5
10–20	8
20–40	16
40–60	7
60–90	4
	40

Range is just the difference between the upper limit of the highest class and the lower limit of the lowest class. So Range is 90 – 0 = 90. For Q.D., first calculate cumulative frequencies as follows:

Class-Intervals CI	*Frequencies f*	*Cumulative Frequencies c. f.*
0–10	5	05
10–20	8	13
20–40	16	29
40–60	7	36
60–90	4	40
	n = 40	

Q_1 is the size of $\frac{n}{4}$th value in a continuous series. Thus it is the size of the 10th value. The class containing the 10th value is 10–20. Hence Q_1 lies in class 10–20. Now, to calculate the exact value of Q_1, the following formula is used:

$$Q_1 = L + \frac{\frac{n}{4} - cf}{f} \times i$$

Where L = 10 (lower limit of the relevant Quartile class)

c.f. = 5 (Value of c.f. for the class preceding the Quartile class)

i = 10 (interval of the Quartile class), and

f = 8 (frequency of the Quartile class) Thus,

$$Q_1 = 10 + \frac{10-5}{8} \times 10 = 16.25$$

Similarly, Q_3 is the size of $\frac{3n}{4}$th

value; i.e., 30th value, which lies in class 40–60. Now using the formula for Q_3, its value can be calculated as follows:

$$Q_3 = L + \frac{\frac{3n}{4} - c.f.}{f} \times i$$

$$Q_3 = 40 + \frac{30 - 29}{7} \times 20$$

$$Q_3 = 42.87$$

$$Q.D. = \frac{42.87 - 16.25}{2} = 13.31$$

In individual and discrete series, Q_1 is the size of $\frac{n+1}{4}$th value, but in a continuous distribution, it is the size of $\frac{n}{4}$th value. Similarly, for Q_3 and median also, n is used in place of n+1.

If the entire group is divided into two equal halves and the median calculated for each half, you will have the median of better students and the median of weak students. These medians differ from the median of the entire group by 13.31 on an average. Similarly, suppose you have data about incomes of people of a town. Median income of all people can be calculated. Now if all people are divided into two equal groups of rich and poor, medians of both groups can be calculated. Quartile Deviation will tell you the average difference between medians of these two groups belonging to rich and poor, from the median of the entire group.

Quartile Deviation can generally be calculated for open-ended distributions and is not unduly affected by extreme values.

3. Measures of Dispersion from Average

Recall that dispersion was defined as the extent to which values differ from their average. Range and Quartile Deviation are not useful in measuring, how far the values are, from their average. Yet, by calculating the spread of values, they do give a good idea about the dispersion. Two measures which are based upon deviation of the values from their average are Mean Deviation and Standard Deviation.

Since the average is a central value, some deviations are positive and some are negative. If these are added as they are, the sum will not reveal anything. In fact, the sum of deviations from Arithmetic Mean is always zero. Look at the following two sets of values.

Set A :	5,	9,	16
Set B :	1,	9,	20

You can see that values in Set B are farther from the average and hence more dispersed than values in Set A. Calculate the deviations from Arithmetic Mean and sum them up. What do you notice? Repeat the same with Median. Can you comment upon the quantum of variation from the calculated values?

© BSTBPC WEB COPY NOT TO BE PUBLISHED

Mean Deviation tries to overcome this problem by ignoring the signs of deviations, i.e., it considers all deviations positive. For standard deviation, the deviations are first squared and averaged and then square root of the average is found. We shall now discuss them separately in detail.

Mean Deviation

Suppose a college is proposed for students of five towns A, B, C, D and E which lie in that order along a road. Distances of towns in kilometres from town A and number of students in these towns are given below:

Town	*Distance from town A*	*No. of Students*
A	0	90
B	2	150
C	6	100
D	14	200
E	18	80
		620

Now, if the college is situated in town A, 150 students from town B will have to travel 2 kilometers each (a total of 300 kilometres) to reach the college. The objective is to find a location so that the average distance travelled by students is minimum.

You may observe that the students will have to travel more, on an average, if the college is situated at town A or E. If on the other hand, it is somewhere in the middle, they are likely to travel less. Mean deviation is the appropriate statistical tool to estimate the average distance travelled by students.Mean Deviation is the arithmetic mean of the differences of the values from their average. The average used is either the arithmetic mean or median.

(Since the mode is not a stable average, it is not used to calculate Mean Deviation.)

Activities

- Calculate the total distance to be travelled by students if the college is situated at town A, at town C, or town E and also if it is exactly half way between A and E.
- Decide where, in your opinion, the college should be established, if there is only one student in each town. Does it change your answer?

Calculation of Mean Deviation from Arithmetic Mean for ungrouped data.

Direct Method

Steps:

(i) The A.M. of the values is calculated

(ii) Difference between each value and the A.M. is calculated. All differences are considered positive. These are denoted as $|d|$

(iii) The A.M. of these differences (called deviations) is the Mean Deviation.

i.e. $\text{M.D.} = \dfrac{\Sigma |d|}{n}$

Example 3

Calculate the Mean Deviation of the following values; 2, 4, 7, 8 and 9.

The A.M. $= \frac{\Sigma X}{n} = 6$

X	*|d|*
2	4
4	2
7	1
8	2
9	3
	12

$$M.D._{(\bar{x})} = \frac{12}{5} = 2.4$$

Assumed Mean Method

Mean Deviation can also be calculated by calculating deviations from an assumed mean. This method is adopted especially when the actual mean is a fractional number. (Take care that the assumed mean is close to the true mean).

For the values in example 3, suppose value 7 is taken as assumed mean, M.D. can be calculated as under:

Example 4

X	*|d|*
2	5
4	3
7	0
8	1
9	2
	11

In such cases, the following formula is used,

$$M.D._{(\bar{x})} = \frac{\Sigma|d| + (\bar{x} - A\bar{x})(\Sigma f_B - \Sigma f_A)}{n}$$

Where $\Sigma |d|$ is the sum of absolute deviations taken from the assumed mean.

$\bar{x}$ is the actual mean.

$A\bar{x}$ is the assumed mean used to calculate deviations.

Σf_B is the number of values below the actual mean including the actual mean.

Σf_A is the number of values above the actual mean.

Substituting the values in the above formula:

$$M.D._{(\bar{x})} = \frac{11 + (6 - 7)(2 - 3)}{5} = \frac{12}{5} = 2.4$$

Mean Deviation from median for ungrouped data.

Direct Method

Using the values in example 3, M.D. from the Median can be calculated as follows,

(i) Calculate the median which is 7.

(ii) Calculate the absolute deviations from median, denote them as |d|.

(iii) Find the average of these absolute deviations. It is the Mean Deviation.

Example 5

	[X-Median]
X	*|d|*
2	5
4	3
7	0
8	1
9	2
	11

© BSTBPC WEB COPY NOT TO BE PUBLISHED

M. D. from Median is thus,

$$M.D._{(median)} = \frac{\Sigma |d|}{n} = \frac{11}{5} = 2.2$$

Short-cut method

To calculate Mean Deviation by short cut method, a value (A) is used to calculate the deviations and the following formula is applied.

$$M.D._{(Median)} = \frac{\Sigma |d| + (Median - A)(\Sigma f_B - \Sigma f_A)}{n}$$

where, A = the constant from which deviations are calculated. (Other notations are the same as given in the assumed mean method).

Mean Deviation from Mean for Continuous distribution

TABLE 6.2

Profits of companies (Rs in lakhs) *Class-intervals*	Number of Companies
10–20	5
20–30	8
30–50	16
50–70	8
70–80	3
	40

Steps:

(i) Calculate the mean of the distribution.

(ii) Calculate the absolute deviations $|d|$ of the class midpoints from the mean.

(iii) Multiply each $|d|$ value with its corresponding frequency to get $f|d|$ values. Sum them up to get $\Sigma f|d|$.

(iv) Apply the following formula,

$$M.D._{(\bar{x})} = \frac{\Sigma f |d|}{\Sigma f}$$

Mean Deviation of the distribution in Table 6.2 can be calculated as follows:

Example 6

C.I.	*f*	*m.p.*	*|d|*	*f|d|*
10–20	5	15	25.5	127.5
20–30	8	25	15.5	124.0
30–50	16	40	0.5	8.0
50–70	8	60	19.5	156.0
70–80	3	75	34.5	103.5
	40			*519.0*

$$M.D._{(\bar{x})} = \frac{\Sigma f |d|}{\Sigma f} = \frac{519}{40} = 12.975$$

Mean Deviation from Median

TABLE 6.3

Class intervals	*Frequencies*
20–30	5
30–40	10
40–60	20
60–80	9
80–90	6
	50

The procedure to calculate Mean Deviation from the median is the same as it is in case of M.D. from Mean, except that deviations are to be taken from the median as given below:

Example 7

C.I.	*f*	*m.p.*	*/d/*	*f/d/*
20–30	5	25	25	125
30–40	10	35	15	150
40–60	20	50	0	0
60–80	9	70	20	180
80–90	6	85	35	210
	50			*665*

$$M.D._{(Median)} = \frac{\Sigma f \mid d \mid}{\Sigma f}$$

$$= \frac{665}{50} = 13.3$$

Mean Deviation: Comments

Mean Deviation is based on all values. A change in even one value will affect it. It is the least when calculated from the median i.e., it will be higher if calculated from the mean. However it ignores the signs of deviations and cannot be calculated for open-ended distributions.

Standard Deviation

Standard Deviation is the positive square root of the mean of squared deviations from mean. So if there are five values x_1, x_2, x_3, x_4 and x_5, first their mean is calculated. Then deviations of the values from mean are calculated. These deviations are then squared. The mean of these squared deviations is the variance. Positive square root of the *variance* is the standard deviation.

(Note that Standard Deviation is calculated on the basis of the mean only).

Calculation of Standard Deviation for ungrouped data

Four alternative methods are available for the calculation of standard deviation of individual values. All these methods result in the same value of standard deviation. These are:

(i) Actual Mean Method
(ii) Assumed Mean Method
(iii) Direct Method
(iv) Step-Deviation Method

Actual Mean Method:

Suppose you have to calculate the standard deviation of the following values:

5, 10, 25, 30, 50

Example 8

X	*d (x-$\bar{x}$)*	*d^2*
5	–19	361
10	–14	196
25	+1	1
30	+6	36
50	+26	676
	0	*1270*

Following formula is used:

$$\sigma = \sqrt{\frac{\Sigma d^2}{n}}$$

$$\sigma = \sqrt{\frac{1270}{5}} = \sqrt{254} = 15.937$$

Do you notice the value from which deviations have been calculated in the above example? Is it the Actual Mean?

Assumed Mean Method

For the same values, deviations may be calculated from any arbitrary value

© PSEB WEB COPY NOT TO BE PUBLISHED

$A_{\bar{x}}$ such that $d = X - A_{\bar{x}}$. Taking $A_{\bar{x}} = 25$, the computation of the standard deviation is shown below:

Example 9

X	*d (x-A$\bar{x}$)*	*d^2*
5	–20	400
10	–15	225
25	0	0
30	+5	25
50	+25	625
	–5	*1275*

Formula for Standard Deviation

$$\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$$

$$\sigma = \sqrt{\frac{1275}{5} - \left(\frac{-5}{5}\right)^2} = \sqrt{254} = 15.937$$

Note that the sum of deviations from a value other than actual mean will not be equal to zero.

Direct Method

Standard Deviation can also be calculated from the values directly, i.e., without taking deviations, as shown below:

Example 10

X	*X^2*
5	25
10	100
25	625
30	900
50	2500
120	*4150*

(This amounts to taking deviations from zero)

Following formula is used.

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

or $$\sigma = \sqrt{\frac{4150}{5} - (24)^2}$$

or $$\sigma = \sqrt{254} = 15.937$$

Standard Deviation is not affected by the value of the constant from which deviations are calculated. The value of the constant does not figure in the standard deviation formula. Thus, Standard Deviation is *Independent of Origin.*

Step-deviation Method

If the values are divisible by a common factor, they can be so divided and standard deviation can be calculated from the resultant values as follows:

Example 11

Since all the five values are divisible by a common factor 5, we divide and get the following values:

x	*x'*	*d (x'-$\bar{x}$')*	*d^2*
5	1	–3.8	14.44
10	2	–2.8	7.84
25	5	+0.2	0.04
30	6	+1.2	1.44
50	10	+5.2	27.04
		0	*50.80*

(Steps in the calculation are same as in actual mean method).

The following formula is used to calculate standard deviation:

$$\sigma = \sqrt{\frac{\Sigma d^2}{n}} \times c$$

$$x' = \frac{x}{c}$$

c = common factor

Substituting the values,

$$\sigma = \sqrt{\frac{50.80}{5}} \times 5$$

$$\sigma = \sqrt{10.16} \times 5$$

$$\sigma = 15.937$$

Alternatively, instead of dividing the values by a common factor, the deviations can be divided by a common factor. Standard Deviation can be calculated as shown below:

Example 12

x	*d (x-25)*	*d' (d/5)*	*d'²*
5	–20	–4	16
10	–15	–3	9
25	0	0	0
30	+5	+1	1
50	+25	+5	25
		–1	*51*

Deviations have been calculated from an arbitrary value 25. Common factor of 5 has been used to divide deviations.

$$\sigma = \sqrt{\frac{\Sigma d'^2}{n} - \left(\frac{\Sigma d'}{n}\right)^2} \times c$$

$$\sigma = \sqrt{\frac{51}{5} - \left(\frac{-1}{5}\right)^2} \times 5$$

$$\sigma = \sqrt{10.16} \times 5 = 15.937$$

> Standard Deviation is *not independent of scale.* Thus, if the values or deviations are divided by a common factor, the value of the common factor is used in the formula to get the value of Standard Deviation.

Standard Deviation in Continuous frequency distribution:

Like ungrouped data, S.D. can be calculated for grouped data by any of the following methods:

(i) Actual Mean Method
(ii) Assumed Mean Method
(iii) Step-Deviation Method

Actual Mean Method

For the values in Table 6.2, Standard Deviation can be calculated as follows:

Example 13

(1) CI	*(2) f*	*(3) m*	*(4) fm*	*(5) d*	*(6) fd*	*(7) fd²*
10–20	5	15	75	–25.5	–127.5	3251.25
20–30	8	25	200	–15.5	–124.0	1922.00
30–50	16	40	640	–0.5	–8.0	4.00
50–70	8	60	480	+19.5	+156.0	3042.00
70–80	3	75	225	+34.5	+103.5	3570.75
	40		*1620*		*0*	*11790.00*

Following steps are required:

1. Calculate the mean of the distribution.

$$\bar{x} = \frac{\Sigma fm}{\Sigma f} = \frac{1620}{40} = 40.5$$

2. Calculate deviations of mid-values from the mean so that

$$d = m - \bar{x} \text{ (Col. 5)}$$

3. Multiply the deviations with their

© BSTBU WEB COPY NOT TO BE PUBLISHED

corresponding frequencies to get 'fd' values (col. 6) [Note that $\sum fd = 0$]

4. Calculate 'fd^2' values by multiplying 'fd' values with 'd' values. (Col. 7). Sum up these to get $\sum fd^2$.
5. Apply the formula as under:

$$\sigma = \sqrt{\frac{\Sigma fd^2}{n}} = \sqrt{\frac{11790}{40}} = 17.168$$

Assumed Mean Method

For the values in example 13, standard deviation can be calculated by taking deviations from an assumed mean (say 40) as follows:

Example 14

(1) *CI*	*(2)* *f*	*(3)* *m*	*(4)* *d*	*(5)* *fd*	*(6)* *fd^2*
10–20	5	15	-25	–125	3125
20–30	8	25	-15	–120	1800
30–50	16	40	0	0	0
50–70	8	60	+20	160	3200
70–80	3	75	+35	105	3675
	40			*+20*	*11800*

The following steps are required:

1. Calculate mid-points of classes (Col. 3)
2. Calculate deviations of mid-points from an assumed mean such that $d = m - A\bar{x}$ (Col. 4). Assumed Mean = 40.
3. Multiply values of 'd' with corresponding frequencies to get 'fd' values (Col. 5). (note that the total of this column is not zero since deviations have been taken from assumed mean).
4. Multiply 'fd' values (Col. 5) with 'd' values (col. 4) to get fd^2 values (col. 6). Find $\sum fd^2$.
5. Standard Deviation can be calculated by the following formula.

$$\sigma = \sqrt{\frac{\Sigma fd^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2}$$

$$\text{or } \sigma = \sqrt{\frac{11800}{40} - \left(\frac{20}{40}\right)^2}$$

$$\text{or } \sigma = \sqrt{294.75} = 17.168$$

Step-deviation Method

In case the values of deviations are divisible by a common factor, the calculations can be simplified by the step-deviation method as in the following example.

Example 15

(1) *CI*	*(2)* *f*	*(3)* *m*	*(4)* *d*	*(5)* *d'*	*(6)* *fd'*	*(7)* *fd'^2*
10–20	5	15	–25	–5	–25	125
20–30	8	25	–15	–3	–24	72
30–50	16	40	0	0	0	0
50–70	8	60	+20	+4	+32	128
70–80	3	75	+35	+7	+21	147
	40				*+4*	*472*

Steps required:

1. Calculate class mid-points (Col. 3) and deviations from an arbitrarily chosen value, just like in the assumed mean method. In this example, deviations have been taken from the value 40. (Col. 4)
2. Divide the deviations by a common factor denoted as 'c'. c = 5 in the

above example. The values so obtained are 'd'' values (Col. 5).

3. Multiply 'd'' values with corresponding 'f' values (Col. 2) to obtain 'fd'' values (Col. 6).
4. Multiply 'fd'' values with 'd'' values to get 'fd'^{2}' values (Col. 7)
5. Sum up values in Col. 6 and Col. 7 to get $\sum$ fd' and $\sum$ fd'^{2} values.
6. Apply the following formula.

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{\Sigma f} - \left(\frac{\Sigma fd'}{\Sigma f}\right)^2} \times c$$

or $$\sigma = \sqrt{\frac{472}{40} - \left(\frac{4}{40}\right)^2} \times 5$$

or $$\sigma = \sqrt{11.8 - 0.01} \times 5$$

or $$\sigma = \sqrt{11.79} \times 5$$
$$\sigma = 17.168$$

Standard Deviation: Comments

Standard Deviation, the most widely used measure of dispersion, is based on all values. Therefore a change in even one value affects the value of standard deviation. It is independent of origin but not of scale. It is also useful in certain advanced statistical problems.

5. Absolute and Relative Measures of Dispersion

All the measures, described so far, are absolute measures of dispersion. They calculate a value which, at times, is difficult to interpret. For example, consider the following two data sets:

Set A	500	700	1000
Set B	1,00,000	1,20,000	1,30,000

Suppose the values in Set A are the daily sales recorded by an ice-cream vendor, while Set B has the daily sales of a big departmental store. Range for Set A is 500 whereas for Set B, it is 30,000. The value of Range is much higher in Set B. Can you say that the variation in sales is higher for the departmental store? It can be easily observed that the highest value in Set A is double the smallest value, whereas for the Set B, it is only 30% higher. Thus absolute measures may give misleading ideas about the extent of variation specially when the averages differ significantly.

Another weakness of absolute measures is that they give the answer in the units in which original values are expressed. Consequently, if the values are expressed in kilometers, the dispersion will also be in kilometers. However, if the same values are expressed in meters, an absolute measure will give the answer in meters and the value of dispersion will appear to be 1000 times.

To overcome these problems, relative measures of dispersion can be used. Each absolute measure has a relative counterpart. Thus, for Range, there is Coefficient of Range which is calculated as follows:

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

where L = Largest value
S = Smallest value

Similarly, for Quartile Deviation, it is Coefficient of Quartile Deviation which can be calculated as follows:

© NCERT NOT TO BE REPUBLISHED

Coefficient of Quartile Deviation

$$= \frac{Q_3 - Q_1}{Q_3 + Q_1}$$ where Q_3=3rd Quartile

Q_1 = 1st Quartile

For Mean Deviation, it is Coefficient of Mean Deviation.

Coefficient of Mean Deviation =

$$\frac{\text{M.D.}(\bar{x})}{\bar{x}} \text{ or } \frac{\text{M.D.(Median)}}{\text{Median}}$$

Thus if Mean Deviation is calculated on the basis of the Mean, it is divided by the Mean. If Median is used to calculate Mean Deviation, it is divided by the Median.

For Standard Deviation, the relative measure is called Coefficient of Variation, calculated as below:

Coefficient of Variation

$$= \frac{\text{Standard Deviation}}{\text{Arithmetic Mean}} \times 100$$

It is usually expressed in percentage terms and is the most commonly used relative measure of dispersion. Since relative measures are free from the units in which the values have been expressed, they can be compared even across different groups having different units of measurement.

Example 16

Income Class	*Frequency*	*Percentage Frequency*	*Cumulative Percentage Frequency*	*Mid-points*	*Total Income*	*Percentage Share of Income*	*Cumulative Percentage Share of Income*
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0–5000	5	10.00	10.00	2500	12500	1.29	1.29
5000–10000	10	20.00	30.00	7500	75000	7.71	9.00
10000–20000	18	36.00	66.00	15000	270000	27.76	36.76
20000–40000	10	20.00	86.00	30000	300000	30.85	67.61
40000–50000	7	14.00	100.0	45000	315000	32.39	100.00
Total	50	100.00	-	-	972500	100.00	-

Column (3) = [col. (2) ÷ total of col. (2)] × 100.
Column (6) = [co I. (2) × col. (5)].
Column (7) = [col. (6) ÷ total of col. (6)] × 100.

7. LORENZ CURVE

The measures of dispersion discussed so far give a numerical value of dispersion. A graphical measure called Lorenz Curve is available for estimating inequalities in distribution. You may have heard of statements like 'top 10% of the people of a country earn 50% of the national income while top 20% account for 80%'. An idea about income disparities is given by such figures. Lorenz Curve uses the information expressed in a cumulative manner to indicate the degree of inequality. For example, Lorenz curve of income gives a relationship between percentage of population and its share of income in total income. It is specially useful in comparing the variability of two or more distributions by drawing two or more Lorenz curves on the same axis.

Given below are the monthly incomes of employees of a company.

TABLE 6.4

Incomes	*Number of employees*
0–5,000	5
5,000–10,000	10
10,000–20,000	18
20,000–40,000	10
40,000–50,000	7

Construction of the Lorenz Curve

Since Lorenz curve provides a relationship between percentage of population and its share in cumulative manner, we will first compute percentage population in each income class, and its total, then the cumulative distribution of population. After that we will compute the share, in percentage term, of each income class and then the cumulative distribution of share. Finally, the curve using different combinations of cumulative percentage of population and cumulative percentage of its (population's) share will give the Lorenz curve. Please refer example 16 given above.

Following steps are required.

1. Calculate the percentage frequency by the formula [col. (2) ÷ total of col. (2)]×100 as given in col. (3).
2. Calculate cumulative percentage frequencies as in col. (4).
3. Calculate class mid-points as in col. (5), which can be approximated as the average income of each person in the income class.
4. Now to calculate the share of each income class, we need to know the total income of each class and total income of all classes. The total income of each class will be obtained by multiplying col. (2) with col. (5) as in col. (6).
5. Calculate the percentage share by the formula [col. (6) ÷ total of col. (6)] × 100 as given in col. (7).
6. Calculate cumulative percentage share as in col. (8).
7. Now, on the graph paper, take the cumulative percentage of the variable (share of incomes [i.e., col. (8)] on Y axis and cumulative percentage of frequencies (number of employees) [i.e., col. (4)] on X- axis, as in figure 6.1. Thus each axis will have values from '0' to '100'. The coordinate (0,0) will imply 0% population has 0% share of income, which is obvious.
8. Draw a line joining Co-ordinate (0,0) with (100,100). This is called the line of equal distribution shown as line 'OC' in figure 6.1.
9. Plot the cumulative percentages share with corresponding cumulative percentages of frequency. Join these points to get the curve OAC.

Studying the Lorenz Curve

OC is called the line of equal distribution, since it would imply a situation like, top 20% people earn 20% of total income and top 60% earn 60% of the total income. The farther the curve OAC from this line, the greater is the inequality present in the distribution. If there are two or more curves on the same axes, the one which is the farthest from line OC has the highest inequality. However, if two Lorenz curves intersect each other, no conclusion can be drawn.

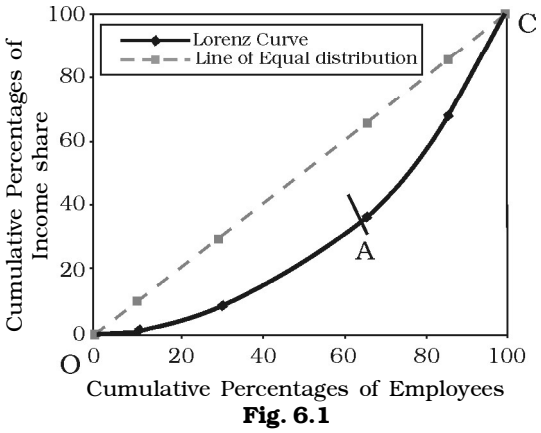


Fig. 6.1

© NCERT NOT TO BE REPUBLISHED

8. Conclusion

Although Range is the simplest to calculate and understand, it is unduly affected by extreme values. QD is not affected by extreme values as it is based on only middle 50% of the data. However, it is more difficult to interpret M.D. and S.D. Both are based upon deviations of values from their average. M.D. calculates average of deviations from the average but ignores signs of deviations and therefore appears to be unmathema-tical. Standard Deviation attempts to calculate average deviation from mean. Like M.D., it is based on all values and is also applied in more advanced statistical problems. It is the most widely used measure of dispersion.

Recap

- A measure of dispersion improves our understanding about the behaviour of an economic variable.
- Range and Quartile Deviation are based upon the spread of values.
- M.D. and S.D. are based upon deviations of values from the average.
- Measures of dispersion could be Absolute or Relative.
- Absolute measures give the answer in the units in which data are expressed.
- Relative measures are free from these units, and consequently can be used to compare different variables.
- A graphic method, which estimates the dispersion from shape of a curve, is called Lorenz Curve.

EXERCISES

1. A measure of dispersion is a good supplement to the central value in understanding a frequency distribution. Comment.
2. Which measure of dispersion is the best and how?
3. Some measures of dispersion depend upon the spread of values whereas some are estimated on the basis of the variation of values from a central value. Do you agree?
4. In a town, 25% of the persons earned more than Rs 45,000 whereas 75% earned more than 18,000. Calculate the absolute and relative values of dispersion.
5. The yield of wheat and rice per acre for 10 districts of a state is as under:

District	1	2	3	4	5	6	7	8	9	10
Wheat	12	10	15	19	21	16	18	9	25	10
Rice	22	29	12	23	18	15	12	34	18	12

Calculate for each crop,
(i) Range
(ii) Q.D.
(iii) Mean Deviation about Mean
(iv) Mean Deviation about Median

(v) Standard Deviation
(vi) Which crop has greater variation?
(vii) Compare the values of different measures for each crop.

6. In the previous question, calculate the relative measures of variation and indicate the value which, in your opinion, is more reliable.
7. A batsman is to be selected for a cricket team. The choice is between X and Y on the basis of their scores in five previous tests which are:

X	25	85	40	80	120
Y	50	70	65	45	80

Which batsman should be selected if we want,
(i) a higher run getter, or
(ii) a more reliable batsman in the team?

8. To check the quality of two brands of lightbulbs, their life in burning hours was estimated as under for 100 bulbs of each brand.

Life (in hrs)	*No. of bulbs*	
	Brand A	*Brand B*
0–50	15	2
50–100	20	8
100–150	18	60
150–200	25	25
200–250	22	5
	100	*100*

(i) Which brand gives higher life?
(ii) Which brand is more dependable?

9. Averge daily wage of 50 workers of a factory was Rs 200 with a Standard Deviation of Rs 40. Each worker is given a raise of Rs 20. What is the new average daily wage and standard deviation? Have the wages become more or less uniform?
10. If in the previous question, each worker is given a hike of 10 % in wages, how are the Mean and Standard Deviation values affected?
11. Calculate the Mean Deviation using mean and Standard Deviation for the following distribution.

Classes	*Frequencies*
20–40	3
40–80	6
80–100	20
100–120	12
120–140	9
	50

12. The sum of 10 values is 100 and the sum of their squares is 1090. Find out the Coefficient of Variation.

© BSTBPC WEBCOPY - NOT TO BE PUBLISHED