Chapter 3: **Measures of Dispersion**

Dispersion

Literal meaning of dispersion is scatter ness. Dispersion is the degree of the scatter ness or deviation of each value in the data set from a measure of central tendency usually the mean or median.

Example

Let the scores of two batches each of size 4, be as follows:

Batch I	49	50	50	51
Batch II	0	0	100	100

The average score for both the batch is 50 that is the mean is same but they are markedly different in their variability or dispersion. This means that although the two sets of data are quite different in nature, the measure of location has failed to bring out this difference.

Different measures of Dispersion

There are several methods of measuring dispersion. These measures can be divided into two groups:

- Absolute measure
- Relative measure

Absolute measures: Absolute measures of variation are expressed in the same statistical unit in which the original data are given such as rupees, kilograms, tones etc. These values may be used to compare the variation in two or more than two distributions provided the variables are expressed in the same units of measurement.

Following are the absolute measures of variation or dispersion

- Range
- Quartile Deviation
- Mean Deviation
- Standard Deviation

Relative measures: The relative measures are described as the ratio of a measure of absolute variation to an average that's why this measure is unit free.

Following are the different relative measures:

- Co-efficient of range
- Co-efficient of quartile deviation
- Co-efficient of mean deviation
- Co-efficient of standard deviation
- Co-efficient of variation

Range

Range is the simplest method of studying variation. It is defined as the difference between the value of the smallest observation and the value of the largest observation included in the distribution. Symbolically,

$$R = L - S$$

Where, $L =$ Largest value and $S =$ Smallest value

The relative measures corresponding to range, called the co-efficient of range, is obtained by applying the following formula

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

In a frequency distribution, range is calculated by taking the difference between the lower limit of the lower class and the upper limit of the highest class.

Example # 1

The following are the points of shares of a company from Monday to Saturday:

Day	Prices (Tk.)
Monday	200
Tuesday	210
Wednesday	208
Thursday	160
Friday	220
Saturday	250

Dr. Mohd. MuziburRahman
Professor, Department of Statistics, JU

Calculate range and co-efficient of range.

Solution: We know that, Range $R = L - S$

Here, $L = 250$ and $S = 160$, $\therefore$ Range = 250-160 = Tk. 90

Co-efficient of Range = $\frac{L-S}{L+S} = \frac{250-160}{250+160}$ = 0.219

Example # 2: Calculate coefficient of range and range from the following data:

Profits (Tk. lakhs)	No. of Companies.
10-20	8
20-30	10
30-40	12
40-50	8
50-60	4

Solution: In a frequency distribution, range is calculated by taking the difference between the lower limit of the lower class and the upper limit of the highest class.

Range = $L - S$ = 60-10 =50

Co-efficient of Range = $\frac{L-S}{L+S} = \frac{60-10}{60+10} = \frac{50}{70}$ = 0.714

Example: The following are the wages of 8 workers of a factory. Find the range of variation and also compute the co-efficient of range.

Wages in Tk.

1400 1450 1520 1380 1485 1495 1575 1440

Solution: Range $= L - S$

Where, $L =$ Largest value and $S =$ Smallest value

Hence, $L = 1575$ and $S = 1380$

$\therefore$ Range $= 1575 - 1380$ =Tk.195

Example: The following are the marks of 80 students of a class. Find the range of variation of marks and also compute co-efficient of range.

Dr. Mohd. MuziburRahman
Professor, Department of Statistics, JU

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No.of Students	4	12	20	18	15	8	2	1

Solution: Range $= L - S$, here, L=80, S=0, $\therefore$ Range = 80-0
= 80

Co-efficient of Range $= \frac{L-S}{L+S} = \frac{80-0}{80+0} = 1$

Merits:

- It is independent of the measure of central tendency and easy to calculate and understand.
- It gives us a quick idea of the variability of the observations involving least amount of time and calculations.

Limitations

- Range is not based on each and every observation of the distribution.
- Range cannot be computed in case of open-end distributions.
- Range cannot tell us extreme observations. For example, observe the following series:

Series A	6	46	46	46	46	46	46	46
Series B	6	6	6	6	6	6	6	46
Series C	6	10	15	25	30	32	40	46

In all the series range is the same (i.e. 46-6 = 40) but it doesn't mean that the distributions are alike. Range is, therefore, more unreliable as a guide to the variation of the values within a distribution.

Uses of Range:

Although range measures the total variation or spread of a set of observation, its uses are prominent in following fields:

Quality control: It is widely used in production process to control the quality of the product without 100% inspection.

Fluctuations in the share prices: Range is useful in studying the variations in the price of stocks and shares and other commodities. They are very sensitive to price changes from one period to another.

Dr. Mohd. MuzibutRahman
Professor, Department of Statistics, JU

Weather forecasts: The meteorological department uses the range to determine the difference between the minimum and maximum temperature, which is a very useful index for people to know the limits of temperature in a particular season. Also, maximum and minimum values of other climate factors such as rainfall, humidity, wind velocity etc. are very important from the metrological point of view.

Quartile deviation or Semi interquartile range: The limitations or the disadvantages of the range can partially be overcome by using another measure of variation which measures the spread over the middle half of the values in the data set to minimize the influence of outlier. Quartile deviation is defined as: $QD = \frac{Q_3 - Q_1}{2}$ and the coefficient of quartile deviation is defined as: Coefficient of $QD = \frac{Q_3 - Q_1}{Q_3 + Q_1}$.

Problem: Following distribution shows the number of defective items of a lot (containing 100 items) of 120 lots.

No. of defective item	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70
No. of lots	3	7	11	22	40	24	9	4

Compute the range within which the middle 50 percent of the defective lots belongs to.

Merits of QD: (i) It is not difficult to calculate but can only be used to evaluate the variation of observed values within the middle of the data. Its value is not affected by the extreme values in the data set.
(ii) It is an appropriate measure of variation for a data set summarized in open-end class interval.
(iii) Since it is a positional measure of variation, therefore it is useful in case of erratic or highly skewed distribution.

Uses: QD is used to measure the variation income, wealth and land distribution.

Mean deviation:

Mean deviation is obtained by calculating the absolute deviations of each observation from mean (or median or mode) and then averaging these deviations by taking their arithmetic mean.

Let $X_1, X_2, ..., X_n$ be n observations of a variable with mean ($\bar{x}$), median (M_e) and Mode (M_o) then mean deviation is defined by:

Dr. Mohd. MuziburRahman
Professor, Department of Statistics, JU

For Ungrouped Data

$$M.D_{(\bar{x})} = \frac{1}{n}\sum\left|x-\bar{x}\right|$$

$$M.D_{(M_e)} = \frac{1}{n}\sum\left|x-M_e\right|$$

$$M.D_{(M_o)} = \frac{1}{n}\sum\left|x-M_o\right|$$

For grouped Data

$$M.D_{(\bar{x})} = \frac{1}{n}\sum f\left|x-\bar{x}\right|$$

$$M.D_{(M_e)} = \frac{1}{n}\sum f\left|x-M_e\right|$$

$$M.D_{(M_o)} = \frac{1}{n}\sum f\left|x-M_o\right|$$

Where, $n = \sum f$

Co-efficient of Mean deviation (C.M.D)

Co-efficient of mean deviation is the ratio of the mean deviation measured from certain measure of central location to the corresponding measure of central location and is defined as follows:

$$C.M.D_{(\bar{x})} = \frac{M.D_{\bar{x}}}{\bar{x}}$$

$$C.M.D_{(M_e)} = \frac{M.D_{(M_e)}}{M_e}$$

$$C.M.D_{(M_o)} = \frac{M.D_{(M_o)}}{M_o}$$

Problem: Calculate the mean deviation from

(a) Arithmetic mean
(b) Mode
(c) Median

Dr. Mohd. MuziburRahman
Professor, Department of Statistics, JU

In respect of the marks obtained by nine students given below and show that the mean deviation from median is minimum.

Marks (Out of 25)	7	4	10	9	15	12	7	9	7

Solution: Calculation of mean deviation from mean, median and mode:

We know that,

$$M.D_{(\bar{x})} = \frac{1}{n}\sum\left|x-\bar{x}\right|$$

$$M.D_{(M_e)} = \frac{1}{n}\sum\left|x-M_e\right|$$

$$M.D_{(M_o)} = \frac{1}{n}\sum\left|x-M_o\right|$$

Now, Mean $\bar{x} = \frac{\sum x_i}{n} = \frac{80}{9} = 8.89$

For calculating median the items have to be arranged

Marks (Out of 25)	4	7	7	7	9	10	10	12	15

$\therefore$ Median = size of $\frac{n+1}{2}$ th item = $\frac{9+1}{2}$ = 5th item

Here, Size of 5^{th} item = 9.

Hence, median=9.

Mode = 7 (since 7 is repeated the maximum number of items i.e. 3)

Dr. Mohd. MuziburRahman
Professor, Department of Statistics, JU

Calculation of Mean Deviation

Marks x	Deviations from Mean = $\lvert x_i - \bar{x} \rvert$	Deviations from Median = $\lvert x_i - Me \rvert$	Deviations from Mode = $\lvert x_i - Mo \rvert$
7	1.89	2	0
4	4.89	5	3
10	1.11	1	3
9	0.11	0	2
15	6.11	6	8
12	3.11	3	5
7	1.89	2	0
9	0.11	0	2
7	1.89	2	0
Total	$\sum \lvert x_i - \bar{x} \rvert = 21.11$	$\sum \lvert x_i - M_e \rvert = 21$	$\sum \lvert x_i - M_o \rvert = 23$

Now we have,

$$M.D_{(\bar{x})} = \frac{1}{n} \sum \lvert x - \bar{x} \rvert = \frac{1}{9} \times 21.11 = 2.34$$

$$M.D_{(M_e)} = \frac{1}{n} \sum \lvert x - M_e \rvert = \frac{1}{9} \times 21 = 2.33$$

$$M.D_{(M_o)} = \frac{1}{n} \sum \lvert x - M_o \rvert = \frac{1}{9} \times 23 = 2.56$$

From these calculations it is clear that the mean deviation is least from median.

For grouped Data

$$M.D_{(\bar{x})} = \frac{1}{n} \sum f \lvert x - \bar{x} \rvert$$

$$M.D_{(M_e)} = \frac{1}{n} \sum f \lvert x - M_e \rvert$$

$$M.D_{(M_o)} = \frac{1}{n} \sum f \lvert x - M_o \rvert$$

Dr. Mohd. MuziburRahman
Professor, Department of Statistics, JU

Where, $n = \sum f$

Problem: Calculate mean deviation (taking deviations from mean) from the following data:

x	2	4	6	8	10
f	1	4	6	4	1

Solution

We know that,

$$M.D_{(\bar{x})} = \frac{1}{n}\sum f\left|x-\bar{x}\right|$$

Now,

$$\bar{x} = \frac{\sum fx}{n}$$

By using calculator we get, $\sum fx = 96$ and $n = \sum f = 16$.

$$\therefore \bar{x} = \frac{\sum fx}{n} = \frac{96}{16} = 6$$

Calculation of mean deviation

x	f	**Deviations from Mean** $=\left\|x-\bar{x}\right\|$	$f\left\|x-\bar{x}\right\|$
2	1	4	4
4	4	2	8
6	6	0	0
8	4	2	8
10	1	4	4
Total			$\sum f\left\|x-\bar{x}\right\| = 24$

$$\therefore M.D_{(\bar{x})} = \frac{1}{n}\sum f\left|x-\bar{x}\right|$$

$$= \frac{1}{16}\times 24$$

$$= 1.5$$

Dr. Mohd. MuziburRahman
Professor, Department of Statistics, JU

Problem: Calculate mean deviation (from mean) from the following data:

Size of the item	3-4	4-5	5-6	6-7	7-8	8-9	9-10
Frequency	3	7	22	60	85	32	8

Solution: We know that,

$$M.D_{(\bar{x})} = \frac{1}{n}\sum f\left|x-\bar{x}\right|$$

Calculation of mean deviation from mean

Size of the item	**Frequency** f	**Mid point** x	**Deviations from Mean** $=\left\|x-\bar{x}\right\|$	$f\left\|x-\bar{x}\right\|$
3-4	3	3.5	3.59	10.77
4-5	7	4.5	2.59	18.13
5-6	22	5.5	1.59	34.98
6-7	60	6.5	0.59	35.4
7-8	85	7.5	0.41	34.85
8-9	32	8.5	1.41	45.12
9-10	8	9.5	2.41	19.28
Total	$n=\sum f = 217$			$\sum f\left\|x_i-\bar{x}\right\|$ **=198.53**

Now we have,

$$\bar{x} = \frac{\sum fx}{n} = \frac{1538.5}{217} = 7.09$$

$$\therefore\ M.D_{(\bar{x})} = \frac{1}{n}\sum f\left|x-\bar{x}\right|$$

Dr. Mohd. MuziburRahman
Professor, Department of Statistics, JU

$$= \frac{1}{217} \times 198.53 = 0.9148 = 0.915$$

Problem: Age distribution of hundred life insurance policy holders is as follows:

Age as on nearest birthday	Number
17-20	9
20-26	16
26-36	12
36-41	26
41-51	14
51-56	12
56-61	6
61-71	5

Calculate mean deviation from median age.

Solution: We know that,

$$M.D_{(M_e)} = \frac{1}{n} \sum f \left| x - M_e \right|$$

Calculation of mean deviation

Age as on nearest birthday	**Number** f	$c.f$	**Mid point** x	**Deviations from Median** $= \left\| x - M_e \right\|$	$f \left\| x - M_e \right\|$
17-20	9	9	18.5	19.75	177.75
20-26	16	25	23	15.25	244.00
26-36	12	37	31	7.25	87.00
36-41	26	63	38.5	0.25	6.50
41-51	14	77	46	7.75	108.50
51-56	12	89	53.5	15.25	183.00
56-61	6	95	58.5	20.25	121.50
61-71	5	100	66	27.75	138.75
Total	$\sum f_i = 100$				$\sum f \left\| x - M_e \right\| = 1067$

Dr. Mohd. MuziburRahman
Professor, Department of Statistics, JU

We know that,

$$\text{Median} = l + \frac{\frac{n}{2} - p.c.f}{f} \times i$$

Here, $\frac{n}{2} = \frac{100}{2} = 50$

Here median class is 36-41.

$$\therefore l = 36,\ p.c.f = 37,\ f = 26,\ i = 5$$

So we have,

$$\text{Median} = 36 + \frac{50-37}{26} \times 5 = 38.25$$

$$\therefore M.D_{(M_e)} = \frac{1}{n}\sum f\left|x - M_e\right| = \frac{1}{100} \times 1067 = 10.67$$

Merits

- It is easy to understand.
- It is relatively easy to calculate.
- It takes all the observations into account.
- It is less affected by the extreme values.

Demerits

- It is not amenable to further algebraic treatment.
- It cannot be calculated if the extreme classes of the frequency distribution are open.
- It is less stable than standard deviation.

Uses

Because of its simplicity in meaning and computation it is specially effective in reports presented to the general public or to groups not familiar with statistical methods. This measure is useful for small samples with no elaborately analysis required. Incidentally it may be mentioned that the National Bureau of Economic Research has found in its work on forecasting business cycles, that the mean deviation in the most practical measure of variation to use for this purpose.

Standard deviation

Dr. Mohd. MuziburRahman
Professor, Department of Statistics, JU

Standard deviation may be defined as the positive square root of the arithmetic mean of the squares of deviations of given observations from their arithmetic mean.

For ungrouped data

Let $x_1, x_2, ..., x_n$ denote n values of a variable x. The standard deviation denoted by S_x is defined as

$$S_x = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} = \sqrt{\frac{1}{n-1}\left\{\sum x^2 - \frac{(\sum x)^2}{n}\right\}}$$

Problem: Find the standard deviation from the weekly wages of ten workers working in a factory:

Workers	wages	workers	wages
A	320	F	340
B	310	G	325
C	315	H	321
D	322	I	320
E	326	J	331

Solution: We know that,

$$S_x = \sqrt{\frac{1}{n-1}\left\{\sum x^2 - \frac{(\sum x)^2}{n}\right\}}$$

By using calculator we get,

$$\sum x^2 = 1043912,\ \sum x = 3230 \text{ and } n = 10 .$$

So that,

$$S_x = \sqrt{\left\{1043912 - \frac{(3230)^2}{10}\right\}} = 8.31$$

For grouped data

If $x_1, x_2, ..., x_n$ occurs with frequency $f_1, f_2, ..., f_n$ respectively, the standard deviation is defined as

Dr. Mohd. MuziburRahman
Professor, Department of Statistics, JU

$$S_x = \sqrt{\frac{\sum f\left(x-\bar{x}\right)^2}{n-1}} = \sqrt{\frac{1}{n-1}\left\{\sum fx^2 - \frac{(\sum fx)^2}{n}\right\}}$$

Problem: An analysis of production rejects resulted in the following figures:

No. of rejects per operator	No. of operators
21-25	5
26-30	15
31-35	28
36-40	42
41-45	15
46-50	12
51-55	3

Solution: We know that

$$S_x = \sqrt{\frac{1}{n-1}\left\{\sum fx^2 - \frac{(\sum fx)^2}{n}\right\}}$$

Where, $n = \sum f$.

Calculation of Standard Deviation

No. of rejects per operator	No. of operators f	Mid point x	x^2	fx	fx^2
21-25	5	23	529	115	2645
26-30	15	28	784	420	11760
31-35	28	33	1089	924	30492
36-40	42	38	1444	1596	60648
41-45	15	43	1849	645	27735
46-50	12	48	2304	576	27648
51-55	3	53	2809	159	8427

Dr. Mohd. MuziburRahman
Professor, Department of Statistics, JU

Total	120		10808	4435	169355

From the data by using calculator we get,

$$\sum fx = 4435,\ n = \sum f = 120 \text{ and } \sum fx^2 = 169355$$

$$\therefore S_x = \sqrt{\frac{1}{119}\left\{\frac{169355}{120} - \left(\frac{4435}{120}\right)^2\right\}}$$

= 6.7642
= 6.76

Merits and Limitations of Standard Deviation

Merits

- The standard deviation is the best measure of variation because of its mathematical characteristics. It is based on every item the distribution.
- It is possible to calculate the combined standard deviation of two or more groups. This is not possible with any other measure.
- For comparing the variability of two or more distributions coefficient of variation is considered to be most appropriate and this measure is based on mean and standard deviation.
- Standard deviation is most prominently used in further statistical work. For example, in comparing skewness, correlation etc., and use in made of standard deviation.

Limitation

- As compared to other measures it is difficult to compute.
- It gives more weight to extreme values and less to those which are near the mean.

Co-efficient of Standard Deviation

Co-efficient of standard deviation is defined by

Dr. Mohd. MuziburRahman
Professor, Department of Statistics, JU

$$C.S.D = \frac{S.D}{Mean} = \frac{\sigma}{\mu} \text{ (for population) and } \frac{s}{\overline{x}} \text{ for sample.}$$

Co-efficient of Variation

The relative measure of dispersion based upon standard deviation is called co-efficient of standard deviation. The co-efficient of standard deviation multiplied by 100 gives the co-efficient of variation.

Thus

Co-efficient of variation (C.V)

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

$$= \frac{\sigma}{\mu} \times 100 \text{ (for population) and } \frac{s}{\overline{x}} \times 100 \text{ (for sample)}$$

Where, σ and μ are both measured in the same units.

Note: Co-efficient variation is more useful when the two distributions are entirely different and the units of measurement are also different. Co-efficient of variation being a pure number is independent of the units of measurement and thus is suitable for comparing the variability, homogeneity and uniformly of two or more distributions. The series having greater C.V is said to be more variable than the other and the series having lesser C.V is said to be more consistent (or homogenous) than the other. Co-efficient of variation is, however unreliable if $\overline{x}$ is near to zero.

Example: Calculate co-efficient of variation from the following data:

Profits (Rs. crores)	10-20	20-30	30-40	40-50	50-60
No. of Companies	8	12	20	6	4

Solution: We know that,

$$C.V = \frac{s}{\overline{x}} \times 100 \text{ and } \overline{x} = \frac{\sum fx}{n}$$

Calculation of coefficient of variation

Profits (Tk. crores)	**Midpoint** (x)	**No. of cost** (f)	x^2	fx	fx^2

Dr. Mohd. MuziburRahman
Professor, Department of Statistics, JU

10-20	15	8	225	120	1800
20-30	25	12	625	300	7500
30-40	35	20	1225	700	24500
40-50	45	6	2025	270	12150
50-60	55	4	3025	220	12100
Total		50	7125	1610	58050

By using calculator we get

$$\sum fx = 1610, \quad \sum fx^2 = 58050 \text{ and } n = \sum f = 50.$$

$$\therefore \bar{x} = \frac{\sum fx}{n} = \frac{1610}{50} = 32.2$$

$$\therefore s = \sqrt{\frac{1}{n-1}\left\{\sum fx^2 - \left(\sum fx\right)^2 / n\right\}} = 11.14$$

$$\therefore \text{Co-efficient of variation } = \frac{s}{\bar{x}} \times 100 = \frac{11.14}{32.2} \times 100 = 34.596 = 34.6\%$$

Example: A purchasing agent obtained samples of 60 watt bulbs from two companies. He had the samples tested in his own laboratory for length of life with the following result:

Length of life (in hours)	Company A	Company B
1700-1900	10	3
1900-2100	16	40
2100-2300	20	12
2300-2500	8	3
2500-2700	6	2

a) Which company's bulbs do you think are better?
b) If prices of both types are the same, which company's bulbs would you buy?

Solution

Calculation of Mean and co-efficient of variation

Length of life (in hours)	Midpoint	Company A	Company B
1700-1900	1800	10	3
1900-2100	2000	16	40

Dr. Mohd. MuziburRahman
Professor, Department of Statistics, JU

2100-2300	2200	20	12
2300-2500	2400	8	3
2500-2700	2600	6	2

For Company A:

By using calculator we get

$$\sum fx = 128800,\ \sum fx^2 = 279840000 \text{ and } n = \sum f = 60$$

$$\therefore \bar{x} = \frac{\sum fx}{n} = \frac{128800}{60} = 2146.67 \text{ and}$$

S= 236.267 = 236.27

$$\therefore \text{Co-efficient of variation } = \frac{S}{\bar{x}} \times 100 = \frac{236.27}{2146.67} \times 100 = 11.00 = 11\%$$

For Company B:

By using calculator we get

$$\sum fx = 124200,\ \sum fx^2 = 258600000 \text{ and } n = \sum f = 60$$

$$\therefore \bar{x} = \frac{\sum fx}{n} = \frac{124200}{60} = 2070 \text{ and}$$

S= 158.429 = 158.43

$$\therefore \text{Co-efficient of variation} = \frac{S}{\bar{x}} \times 100 = \frac{158.43}{2070} \times 100 = 7.65\%$$

(a) Since average is higher in case of company A, hence bulbs of company A are better.

(b) Co-efficient of variation is less for company B. Hence if prices are same we will prefer to buy Company's B's bulbs.

Example: The prices of a Tea company shares in Dhaka and Chittagong Markets during the least ten months are recorded below:

Dr. Mohd. MuziburRahman
Professor, Department of Statistics, JU

Month	Dhaka	Chittagong
January	105	108
February	120	117
March	115	120
April	118	130
May	130	100
June	127	125
July	109	125
August	110	120
September	104	110
October	112	135

Determine the arithmetic mean and standard deviation of the prices of shares. In which market are the shares prices stable?

Solution

For determining in which market prices of shares are more stable we shall compare the co-efficient of variation. Let prices in Dhaka and Chittagong are denoted by X and Y respectively:

Dhaka:

$$\overline{X} = \frac{\sum X}{n} = \frac{1150}{10} = 115$$

S=8.33

$$C.V = \frac{S}{\overline{X}} \times 100 = \frac{8.33}{115} \times 100 = 7.24\%$$

Chittagong:

$$\overline{X} = \frac{\sum X}{n} = \frac{1190}{10} = 119$$

S=10.089 = 10.09

$$C.V = \frac{S}{\overline{X}} \times 100 = \frac{10.09}{119} \times 100 = 8.478 = 8.48\%$$

Dr. Mohd. MuziburRahman
Professor, Department of Statistics, JU

Since the co-efficient of variation is less in Dhaka, hence the share price in the Dhaka market shows greater stability.

Dr. Mohd. MuziburRahman
Professor, Department of Statistics, JU