MEASURES OF DISPERSION

Introduction :

The measure of central tendency serve to locate the center of the distribution, but they do not reveal how the items are spread out on either side of the center. This characteristic of a frequency distribution is commonly referred to as dispersion. In a series all the items are not equal. There is difference or variation among the values. The degree of variation is evaluated by various measures of dispersion. Small dispersion indicates high uniformity of the items, while large dispersion indicates less uniformity. For example consider the following marks of two students.

Student I	Student II
68	85
75	90
65	80
67	25
70	65

Both have got a total of 345 and an average of 69 each. The fact is that the second student has failed in one paper. When the averages alone are considered, the two students are equal. But first student has less variation than second student. Less variation is a desirable characteristic.

Characteristics of a good measure of dispersion:

An ideal measure of dispersion is expected to possess the following properties

1.It should be rigidly defined

2. It should be based on all the items.
3. It should not be unduly affected by extreme items.
4. It should lend itself for algebraic manipulation.
5. It should be simple to understand and easy to calculate

Absolute and Relative Measures :

There are two kinds of measures of dispersion, namely

1.Absolute measure of dispersion

2.Relative measure of dispersion.

Absolute measure of dispersion indicates the amount of variation in a set of values in terms of units of observations. For example, when rainfalls on different days are available in mm, any absolute measure of dispersion gives the variation in rainfall in mm. On the other hand relative measures of dispersion are free from the units of measurements of the observations. They are pure numbers. They are used to compare the variation in two or more sets, which are having different units of measurements of observations. The various absolute and relative measures of dispersion are listed below.

Absolute measure	**Relative measure**
1. Range	1.Co-efficient of Range
2.Quartile deviation	2.Co-efficient of Quartile deviation
3.Mean deviation	3. Co-efficient of Mean deviation
4.Standard deviation	4.Co-efficient of variation

Range and coefficient of Range:

Range:

This is the simplest possible measure of dispersion and is defined as the difference between the largest and smallest values of the variable.

In symbols, Range = L – S. Where L = Largest value, S = Smallest value.
In individual observations and discrete series, L and S are easily identified. In continuous series, the following two methods are followed.

Method 1:
L = Upper boundary of the highest class
S = Lower boundary of the lowest class.

Method 2:
L = Mid value of the highest class.
S = Mid value of the lowest class.

Co-efficient of Range :

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

Example: Find the value of range and its co-efficient for the following data.
7, 9, 6, 8, 11, 10, 4

Solution: L=11, S = 4.
Range = L – S = 11- 4 = 7

$$\text{Co-efficient of Range} = \frac{L-S}{L+S}$$
$$= \frac{11-4}{11+4}$$
$$= \frac{7}{15} = 0.4667$$

Example: Calculate range and its co efficient from the following distribution.

Size:	60-63	63-66	66-69	69-72	72-75
Number:	5	18	42	27	8

Solution: L = Upper boundary of the highest class = 75
S = Lower boundary of the lowest class = 60
Range = L – S = 75 – 60 = 15

$$\text{Co-efficient of Range} = \frac{L-S}{L+S}$$
$$= \frac{75-60}{75+60}$$
$$= \frac{15}{135} = 0.1111$$

Merits and Demerits of Range :

Merits:
1. It is simple to understand.
2. It is easy to calculate.
3. In certain types of problems like quality control, weather forecasts, share price analysis, etc., range is most widely used.

Demerits:

1. It is very much affected by the extreme items.
2. It is based on only two extreme observations.
3. It cannot be calculated from open-end class intervals.
4. It is not suitable for mathematical treatment.
5. It is a very rarely used measure.

Quartile Deviation and Co efficient of Quartile Deviation :

Quartile Deviation (Q.D) :

Definition: Quartile Deviation is half of the difference between the first and third quartiles. Hence, it is called Semi Inter Quartile Range. In Symbols,

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

Among the quartiles Q_1, Q_2 and Q_3, the range $Q_3 - Q_1$ is called inter quartile range and

$\frac{Q_3 - Q_1}{2}$, Semi inter quartile range.

Co-efficient of Quartile Deviation :

$$\text{Co-efficient of } Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Example: Find the Quartile Deviation for the following data:
391, 384, 591, 407, 672, 522, 777, 733, 1490, 2488

Solution: Arrange the given values in ascending order. We get
384, 391, 407, 522, 591, 672, 733, 777, 1490, 2488.
Position of Q_1 is (n+1)/4 = (10+1/)4 = 2.75th item

Q_1 = 2nd value + 0.75 (3rd value – 2nd value)
= 391 + 0.75 (407 – 391)
= 391 + 0.75 × 16
= 391 + 12
= 403

Position Q_3 is 3(n+1)/4 = 3x2.75 = 8.25th item
Q_3 = 8th value + 0.25 (9th value – 8th value)
= 777 + 0.25 (1490 – 777)
= 777 + 0.25 (713)
= 777 + 178.25 = 955.25

Q.D = $(Q_3 - Q_1)/2$= 276.125

Example: Weekly wages of labours are given below. Calculate Q.D and Coefficient of Q.D.

Weekly Wage (Rs.) :	100	200	400	500	600
No. of Weeks :	5	8	21	12	6

Solution:
Position of Q_1 is (N+1)/4 = (52+1/)4 = 13.25th item
Q_1 = 13th value + 0.25 (14th Value – 13th value)
= 13th value + 0.25 (400 – 200)
= 200 + 0.25 (400 – 200)

$= 200 + 0.25\ (200)$

$= 200 + 50 = 250$

Position Q_3 is $3(N+1)/4 = 3 \times 13.25 = 39.75_{th}$ item

$Q_3 = 39_{th}$ value $+ 0.75$ (40_{th} value $- 39_{th}$ value)

$= 500 + 0.75\ (500 - 500)$

$= 500 + 0.75 \times 0$

$= 500$

$Q.D = (Q_3 - Q_1)/2 = (500 - 250)/2 = 125$

$$\text{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{500 - 250}{500 + 250} = \frac{250}{750} = 0.3333$$

Example: For the date given below, give the quartile deviation and coefficient of quartile deviation.

X :	351 – 500	501 – 650	651 – 800	801–950	951–1100
f :	48	189	88	4	28

Solution :

x	f	True class Intervals	Cumulative frequency
351- 500	48	350.5- 500.5	48
501- 650	189	500.5- 650.5	237
651- 800	88	650.5- 800.5	325
801- 950	47	800.5- 950.5	372
951- 1100	28	950.5- 1100.5	400
Total	N = 400		

$$Q_1 = l_1 + \frac{\frac{N}{4} - m_1}{f_1} \times c_1$$

$N/4 = 400/4 = 100$

Q_1 Class is $500.5 - 650.5$, $l_1 = 500.5$, $m_1 = 48$, $f_1 = 189$, $c_1 = 150$

$$\therefore Q_1 = 500.5 + \frac{100 - 48}{189} \times 150$$

$$= 500.5 + \frac{52 \times 150}{189}$$

$$= 500.5 + 41.27$$

$$= 541.77$$

$$Q_3 = l_3 + \frac{3\frac{N}{4} - m_3}{f_3} \times c_3$$

3N/4 = 3 x 100 = 300

Q3 Class is 650.5 – 800.5, l_3 = 650.5, m_3 = 237, f_3 = 88, C_3 = 150

$$\therefore Q_3 = 650.5 + \frac{300 - 237}{88} \times 150$$

$$= 650.5 + \frac{63 \times 150}{88}$$

$$= 650.5 + 107.39$$

$$= 757.89$$

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

$$= \frac{757.89 - 541.77}{2}$$

$$= \frac{216.12}{2}$$

$$= 108.06$$

$$\text{Coefficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$= \frac{757.89 - 541.77}{757.89 + 541.77}$$

$$= \frac{216.12}{1299.66} = 0.1663$$

Mid-range: It is the average of minimum value and maximum value

Mid-

Merits and Demerits of Quartile Deviation

Merits :

1. It is Simple to understand and easy to calculate
2. It is not affected by extreme values.
3. It can be calculated for data with open end classes also.

Demerits:

1. It is not based on all the items. It is based on two positional values Q_1 and Q_3 and ignores the extreme 50% of the items
2. It is not amenable to further mathematical treatment.
3. It is affected by sampling fluctuations.

Standard Deviation and Coefficient of variation:

Standard Deviation :

Karl Pearson introduced the concept of standard deviation in 1893. It is the most important measure of dispersion and is widely used in many statistical formulae. Standard deviation is also called Root-Mean Square Deviation. The reason is that it is the square–root of the mean of the squared deviation from the arithmetic mean. It provides accurate result. Square of standard deviation is called Variance.

Definition:

It is defined as the positive square-root of the arithmetic mean of the Square of the deviations of the given observation from their arithmetic mean. The standard deviation is denoted by the Greek letter σ (sigma)

Calculation of Standard deviation-Individual Series :

There are two methods of calculating Standard deviation in an individual series.

a) Deviations taken from Actual mean

b) Deviation taken from Assumed mean

Example: Calculate the standard deviation from the data 14, 22, 9, 15, 20, 17, 12, 11

Solution: Deviations from actual mean.

Values (X)	$X - \bar{X}$	$(X - \bar{X})^2$
14	-1	1
22	7	49
9	-6	36
15	0	0
20	5	25
17	2	4
12	-3	9
11	-4	16
120		140

$$\bar{X} = \frac{120}{8} = 15$$

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

$$= \sqrt{\frac{140}{8}}$$

$$= \sqrt{17.5} = 4.18$$

Example: The table below gives the marks obtained by 10 students in statistics. Calculate standard deviation.

Student Nos :	1	2	3	4	5	6	7	8	9	10
Marks :	43	48	65	57	31	60	37	48	78	59

Solution: Standard deviation from assumed mean

Nos.	Marks (x)	d=X-A (A=57)	d^2
1	43	-14	196
2	48	-9	81
3	65	8	64
4	57	0	0
5	31	-26	676
6	60	3	9
7	37	-20	400
8	48	-9	81
9	78	21	441
10	59	2	4
n = 10		$\sum d=-44$	$\sum d^2=1952$

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

$$= \sqrt{\frac{1952}{10} - \left(\frac{-44}{10}\right)^2}$$

$$= \sqrt{195.2 - 19.36}$$

$$= \sqrt{175.84} = 13.26$$

Example: calculate the standard deviation from the following data: 4, 6, 8, 14, 18

Solution:

X	x^2
4	16
6	36
8	64
14	6
18	324
$\sum x = 50$	$\sum x^2 = 50$

We know standard deviation(σ) = $\sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$=5.21

For discrete series,

Standard deviation(σ) = $\sqrt{\frac{\sum fx^2}{N} - \left(\frac{\sum fx}{N}\right)^2}$

For Continuous series,

Standard deviation(σ) = $\sqrt{\frac{\Sigma f x^2}{N} - \left(\frac{\sum f x}{N}\right)^2}$, where x is mid value of class interval.

Variance: Square of the standard deviation is Variance. It is denoted by σ^2

Combined Standard Deviation:
If a series of N_1 items has mean $\overline{X}_1$ and standard deviation σ_1 and another series of N_2 items has mean $\overline{X}_2$ and standard deviation σ_2, we can find out the combined mean and combined standard deviation by using the formula.

$$\overline{X}_{12} = \frac{N1\,\overline{X}_1 + N_2\,\overline{X}_2}{N_1 + N_2}$$

and

$$\sigma_{12} = \sqrt{\frac{N_1\,\sigma_1^2 + N_2\,\sigma_2^2 + N_1 d_1^2 + N_2 d_2^2}{N_1 + N_2}}$$

Where $d_1 = \overline{X}_1 - \overline{X}_{12}$

$d_2 = \overline{X}_2 - \overline{X}_{12}$

Merits and Demerits of Standard Deviation:
Merits:
1. It is rigidly defined and its value is always definite and based on all the observations and the actual signs of deviations are used.
2. As it is based on arithmetic mean, it has all the merits of arithmetic mean.
3. It is the most important and widely used measure of dispersion.
4. It is possible for further algebraic treatment.
5. It is less affected by the fluctuations of sampling and hence stable.
6. It is the basis for measuring the coefficient of correlation and sampling.

Demerits:
1. It is not easy to understand and it is difficult to calculate.
2. It gives more weight to extreme values because the values are squared up.
3. As it is an absolute measure of variability, it cannot be used for the purpose of comparison.

Coefficient of Variation :
The Standard deviation is an absolute measure of dispersion. It is expressed in terms of units in which the original figures are collected and stated. The standard deviation of heights of students cannot be compared with the standard deviation of weights of students, as both are expressed in different units, i.e heights in centimeter and weights in kilograms. Therefore the standard deviation must be converted into a relative measure of dispersion for the purpose of comparison. The relative measure is known as the coefficient of variation. The coefficient of variation is obtained by dividing the standard deviation by the mean and multiply it by 100. It is always expressed as percentage.

symbolically,

Coefficient of Variation (C.V.) = $\frac{\sigma}{\bar{x}} \times 100\%$

If we want to compare the variability of two or more series, we can use C.V. The series or groups of data for which the C.V. is greater indicate that the group is more variable, less stable, less uniform, less consistent, less equitable or less homogeneous. If the C.V. is less, it indicates that the group is less variable, more stable, more uniform, more consistent, more equitable or more homogeneous.

Example : In two factories A and B located in the same industrial area, the average weekly wages (in rupees) and the standard deviations are as follows:

Factory	Average	Standard Deviation	No. of workers
A	34.5	5	476
B	28.5	4.5	524

Given $N_1 = 476$, $\overline{X}_1 = 34.5$, $\sigma_1 = 5$; $N_2 = 524$, $\overline{X}_2 = 28.5$, $\sigma_2 = 4.5$

1. Total wages paid by factory A = 34.5 × 476 = Rs.16.422

Total wages paid by factory B = 28.5 × 524 = Rs.14,934.

Therefore factory A pays out larger amount as weekly wages.

2. C.V. of distribution of weekly wages of factory A and B are

Coefficient of Variation (C.V.) for A = $\frac{\sigma}{\bar{x}} \times 100\% = \frac{5}{34.5} \times 100\%$ = 14.49%

Coefficient of Variation (C.V.) for B = $\frac{\sigma}{\bar{x}} \times 100\% = \frac{4.5}{28.5} \times 100\%$ = 15.79%

Factory B has greater variability in individual wages, since C.V. of factory B is greater than C.V of factory A.