1. Measures of Dispersion: How much the observations vary from one to another in a data set is called variation or dispersion. Descriptive measures that indicate the amount of variation in a data set are called measures of dispersion or measures of variation.

2. Types of measures of dispersion: Measures of dispersion are two types,

a) Absolute measures of dispersion
b) Relative measures of dispersion

3. Absolute measures of dispersion: The descriptive measures that only measure the inherent variation of a data set, but they cannot be used directly to compare the variability of two data sets are called absolute measures of dispersion.

The frequently used absolute measures of dispersion are,

a) Range
b) Mean deviation
c) Variance
d) Quartile deviation

4. Range: The range of a data set is the difference between the largest and the smallest observations.

$$Range, R = Largest\ observation - Smallest\ observation$$

Math: Following are the heights of the players of Team 1 and Team 2:

Team 1 (in cm)	142	143	146	148	146
Team 2 (in cm)	137	142	146	154	146

Compute range and interpret.

Solution:

For Team 1, $Range, R = 148 - 142 = 6cm$

For Team 2, $Range, R = 154 - 137 = 17cm$

Interpretation: The difference between the heights of the tallest and the shortest player on Team 1 is 6 cm, whereas the difference for Team 2 is 17 cm.

5. Merits of Range:

a) Range is the simplest though crude measure of dispersion.
b) It is rigidly defined
c) Easiest to compute

6. Demerits of Range:

a) Range is not based on the entire set of data
b) Range is very much affected by fluctuations of sampling
c) Range cannot tell us anything about the character of the distribution within two extreme observations.

7. Mean deviation: Mean deviation of a data set is an average of absolute deviations of the observations from mean (or median). Let $x_1, \dots, x_n$ denote n observations in a data set. Then the mean deviation about mean can be written as,

$$MD = \frac{\sum |x_i - \bar{x}|}{n}$$

Math: Find the mean deviation of data (height in cm) values are, 142, 143, 146, 146, 148.

Solution:

First, we calculate the mean of these observations,

$$\bar{x} = \frac{\sum x_i}{n} = \frac{(142 + 143 + 146 + 146 + 148)}{5} = \frac{725}{5} = 145$$

Now, create a table for mean deviation calculation,

x_i	$\lvert x_i - \bar{x} \rvert$
142	3
143	2
146	1
146	1
148	3
Total	10

$$\therefore MD = \frac{\sum |x_i - \bar{x}|}{n} = \frac{10}{5} = 2$$

Interpretation: Approximately, on average, the players' heights deviate from the mean height of 145 cm by approximately 2 cm.

8. Merits of MD:

a) Mean deviation is rigidly defined and is easy to understand and calculate

b) Mean deviation is based on all the observations and is thus definitely a better measure of dispersion than the range and quartile deviation.

c) Since mean deviation is based on the deviations about an average, it provides a better measure for comparison about the formation of different distributions.

d) it is less affected by extreme observations.

9. Demerits of MD:

a) The strongest objection against mean deviation is that while computing its value we take the absolute value of the deviations about an average and ignore the signs of the deviations.

10. Variance: The variance is the arithmetic mean of squared deviations of the observations from the arithmetic mean.

In our study, we have two types of variances,

- Population variance: Let, $x_1, \dots, x_N$ be N observations in a population and μ be the population mean. Then the population variance denoted by σ^2 is defined as,

$$\sigma^2 = \frac{\sum(x-\mu)^2}{N}$$

For grouped, $\sigma^2 = \frac{\sum f_i(x_i-\mu)^2}{N}$

- Sample variance: Let, $x_1, \dots, x_n$ be n observations in a sample and $\bar{x}$ be the sample mean. Then the population variance denoted by s^2 is defined as,

$$s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$$

For grouped, $s^2 = \frac{\sum f_i(x_i-\bar{x})^2}{n-1}$

11. Standard deviation: The positive square root of variance is called standard deviation.

- Population standard deviation: $\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}}$
- Sample standard deviation: $s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$

Math: Set A: 18, 25, 10, 12. And Set B: 32, 30, 20, 10.

Find population standard deviation and compare the variability between these two data sets based on Standard deviation.

*****Hints: Find Standard deviation for both the data. Then the data set with lower standard deviation will have higher uniformity.***

Math: Calculate the Sample variance and standard deviation for the data set 1, 2, 2, 3, 4, 5.

12. Merits of SD: Among all measures of dispersion Standard Deviation is considered superior because it possesses almost all the requisite characteristics of a good measure of dispersion. It is based on all the observations of the data set. It is amenable to further mathematical calculation.

13. Demerits of SD: It is more affected by extreme values, and as such has not found favor with economists or businessmen who are more interested in the results of the modal class.

14. Relative measures of dispersion: To compare the extent of variation of two or more data sets, whether having different or identical units of measurements, the absolute measures of variation are expressed in the form of coefficients. These coefficients are called relative measures of dispersion. These coefficients are independent on the units of measurements; i.e., they are pure numbers.

These measures are,

a) Coefficient of range
b) Coefficient of mean deviation
c) Coefficient of variation
d) Coefficient of quartile deviation

15. Coefficient of variation: The coefficient of variation is the ratio of the standard deviation to the mean. It is usually expressed as percentage. Mathematically,

- $CV = \frac{\sigma}{\mu} \times 100; For\ population$
- $CV = \frac{s}{\bar{x}} \times 100; For\ Sample$

*** The CV cannot be used when arithmetic mean is zero. If arithmetic mean is negative, the coefficient of variation becomes negative. So, to compare the variability between two distributions, the sign of the coefficient of variation is ignored. The coefficient of variation is meaningful only if the variable is measured on an interval or ratio scale.

16. Purposes of studying coefficient of variation (C.V):

- To compare the variability of two or more data sets.
- To compare the Consistency, Uniformity, Homogeneity, Stability of two or more data sets.
- Since c.v. is a unit free measure, we can compare two or more data sets having different units.

*** When CV is high, we can conclude that the presence of variability is high.
*** When CV is low, we can conclude that the presence of variability is low. Low variability means,

— High uniformity
— High consistency
— High stability
— High homogeneity

Math: The mean final exam marks of Section A is 30 out of 40 with standard deviation 4 and the mean final exam marks of Section B is 25 out of 40 with standard deviation 6. Which Section is more consistent in getting final exam marks?

Solution:

For section A, $\bar{x}_A = 30$, and $SD_A = 4$

$$\therefore CV_A = \frac{SD_A}{\bar{x}_A} = \frac{4}{30} = 13.3\%$$

For section B, $\bar{x}_B = 25$, and $SD_B = 6$

$$\therefore CV_B = \frac{SD_B}{\bar{x}_B} = \frac{6}{25} = 24\%$$

Since, the CV_A is lower than the CV_B. Thus, Section A is more consistent in getting final exam marks.

Math: Population mean of 10 items is 50 and population SD is 14. Find the sum of the squares of all the items.

Solution: Here,

$$\mu = 50, SD = 14, \sum x_i^2 = ?$$

We know that, the population variance is,

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N} = \frac{\sum x_i^2 - N\mu^2}{N}$$

$$\Rightarrow (14)^2 = \frac{\sum x_i^2 - (10 \times (50)^2)}{10}$$

$$\Rightarrow 196 \times 10 = \sum x_i^2 - 25000$$

$$\therefore \sum x_i^2 = 26960$$

$$\boldsymbol{\sum(x_i - \mu)^2 = \sum(x_i^2 - 2\mu x_i + \mu^2)}$$
$$\boldsymbol{= \sum x_i^2 - 2\mu \sum x_i + \sum \mu^2}$$
$$\boldsymbol{= \sum x_i^2 - 2\mu N\mu + N\mu^2; \left[\mu = \frac{\sum x_i}{N}\right]}$$
$$\boldsymbol{= \sum x_i^2 - 2N\mu^2 + N\mu^2}$$
$$\boldsymbol{= \sum x_i^2 - N\mu^2}$$

Modified formula, $\sigma^2 = \frac{\sum x_i^2 - N\mu^2}{N}$

Modified formula, $s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$

17. Objectives or Significance of the Measures of Dispersion.

a) To find out the reliability of an average.

b) To control the variation of the data from the central value.

c) To compare two or more sets of data regarding their variability.

d) To obtain other statistical measures for further analysis of data.

Extra:

1. Ten persons of varying ages were weighed and the following weights in kg were recorded:

$$110, 125, 125, 147, 117, 125, 136, 157, 124, 110$$

Compute mean deviation about mean. **(Ans: 11.44)**

2. Compute mean deviation by using following table:

Age class	25-30	30-35	35-40	40-45	45-50	50-55
$\boldsymbol{f_i}$	3	9	15	12	7	4

3. Compute variance and standard deviation from the data on weight of ten children as follows: $20, 13, 17, 17, 13, 18, 14, 17, 16$ **(Ans: 5.11 $\boldsymbol{kg^2}$ and 2.26 $\boldsymbol{kg}$)**

4. Heights (in cm) of players of a football team: $142, 143, 146, 146, 148$

Use best measures of dispersion with proper interpretation. **(Ans: SD=2.45 cm)**

5. The following is the share price (in BDT) of a company on seven days. Compute Range, Mean Deviation, and Standard Deviation. Interpret the results.

$$1200, 1210, 1208, 1160, 1225, 1240, 1260$$

6. The following is the minimum temperature (in degree Celsius) of several years recorded from Dhaka and Rajshahi division of Bangladesh.

Year	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
Temp. in Dhaka	12.7	13.0	14.0	12.9	12.6	13.3	14.8	13.5
Temp. in Rajshahi	4.2	5.3	4.9	6.0	5.2	5.4	4.0	4.8

a) Compute average minimum temperature of both of the cities.

b) In which city is the temperature more homogeneous.

7. **Factory A:** Average = 2850 BDT, SD = 20 BDT, No. of employees = 30

Factory B: Average = 3000 BDT, SD = 38 BDT, No. of employees = 20

Which factory does pay more consistent salaries?

8. Compute standard deviation for the data on weight of BBA male students from below table,

Weight	35-40	40-45	45-50	50-55	55-0	60-65	65-70	70-75
f_i	2	5	12	16	12	6	4	3

[Ans: $\bar{x} = 54.17kg, SD = 8.32kg$]