LESSON FOUR: MEASURES OF DISPERSION

4.1 Introduction

- Dispersion refers to the degree to which numerical data tends to spread about an average value. It is the extent of the scatteredness of items around a measure of central tendency.
- The measures of dispersion are also referred to as measures of variation or measures of spread.

4.2 Significance of measuring dispersion

- To determine the reliability of an average
- To serve as a basis for the control of the variability
- To compare two or more series with regard to their variability
- To facilitate the use of other statistical measures

4.3 Properties of a good measure of dispersion

It should be: -

- Simple to understand
- Easy to compute
- Rigidly defined
- Based on each and every item in the distribution
- Amenable to further algebraic calculations
- Have sampling stability
- Not be unduly affected by extreme values

NOTE:

The measures of dispersion which are expressed in terms of the original units of the observations are termed as absolute measures. Such measures are not suitable for comparing the variability of two distributions which are not expressed in the same units of measurements. Therefore it is better to use relative measure of dispersion obtained as ratios or percentages and are thus pure numbers independent of the unit of measurement.

4.4 Measures of dispersion

- Range
- Interquartile Range and Quartile Deviation

- Mean deviation
- Standard deviation / Variance

4.4.1 The Range

It is the difference between the smallest value and the largest value of a series

Example

The following are the prices of shares of a company from Monday to Saturday.

Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Price	200	210	208	160	220	250

Calculate the range.

Solution: Range = L – S

= 250 – 160 = 90

NB:

In case of grouped frequency distribution the range is the difference between the upper class boundary of the largest class and the lower class boundary of the smallest class.

Advantages of the Range

- It is the simplest to understand and compute
- It takes the minimum time to calculate the value of the range

Limitations

- It is not based on each and every value of the distribution
- It is subject to fluctuations of considerable magnitude from sample to sample
- It cannot be computed in case of open-ended distributions
- It does not explain or indicate anything about the character of the distribution within the two extreme observations.

Uses of the range

- Quality control
- Fluctuations of prices

- Weather forecast
- Finding the difference between two values e.g. wages earned by different employees.

4.4.2 The Interquartile Range and Quartile Deviation

Interquartile range: it’s the difference between the third quartile and the first quartile

i.e. Interquartile range = $Q_3 – Q_1$

Quartile Deviation: also called the semi-interquartile range. It’s obtained by dividing the interquartile range by 2.

i.e. Q.D = $\frac{Q_3 - Q_1}{2}$ where Q.D = Quartile Deviation

4.4.3 The Mean Deviation

It is the average amount of scatter of the items in the distribution from the mean, median or mode, ignoring the signs of deviation. If $x_1, x_{2,} ..., x_n$ are n observations then the mean deviation about the mean is calculated as;

For ungrouped data: $M.D = \frac{\sum |x - \overline{x}|}{n}$

For grouped data: $M.D = \frac{\sum f\, |x - \overline{x}|}{\sum f}$

Examples

1. Calculate the mean deviation of the following values

   3000, 4000, 4200, 4400, 4600, 4800, 5800

2. Calculate the average deviation from the mean for the following

Sales (thousands)	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No. of days (f)	3	6	11	3	2

Merits of Mean Deviation

1. It is easy to compute and understand
2. It uses all the data
3. It is less affected by the extreme values
4. Since deviations are taken from a central value, comparison about formation of different distributions can easily be made.
5. It shows the significance of an average in the distribution

Demerits

1. Ignores algebraic signs while taking the deviations
2. Cannot be computed for distributions with open-ended class
3. Rarely used in sociological studies

4.4.4 The Variance and Standard Deviation

- The variance of a set of observations is the average squared deviations of the data points from their mean. Variance is the mean square deviation. It is denoted by s^2 for sample data and σ^2 for population data.
- Standard deviation is the square root of the variance. It is denoted by *s* for sample data and σ for population data.

Computing the Variance

- Variance for ungrouped data

$\delta^2 = \frac{\sum\left(x-\bar{x}\right)^2}{n}$, where $\sum\left(x-\bar{x}\right)^2$ = sum of squares of the deviations from arithmetic mean

- Variance for grouped data

$$\delta^2 = \frac{\sum f\left(x-\bar{x}\right)^2}{\sum f}$$

Computing the standard deviation

Standard deviation for ungrouped data

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

Standard deviation for grouped data

$$\sigma = \sqrt{\frac{\sum f (x - \bar{x})^2}{\sum f}}$$

NB: The computations of δ^2 can be simplified by using the following version of the formula

For ungrouped data: $\delta^2 = \frac{\sum x^2}{n} - (\bar{x})^2$

For grouped data: $\delta^2 = \frac{\sum fx^2}{\sum f} - (\bar{x})^2$

Examples

1. Find the standard deviation of the wages of the following ten workers working in a factory

Worker	A	B	C	D	E	F	G	H	I	J
Weekly Sales	1320	1310	1315	1322	1326	1340	1325	1321	1320	1331

2. An analysis of production rejects resulted in the following figures:

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

Calculate the mean and standard deviation

Combined standard deviation

Combined arithmetic mean for two sets of data with arithmetic means $\overline{x_1}, \overline{x_2}$ and the number of observations n_1 n_2 is given by $\bar{X}_{12} = \frac{N_1 \overline{X}_1 + N_2 \overline{X_2}}{N_1 + N_2}$

Combined standard deviation of two series is given by

$$\delta_{12} = \sqrt{\frac{N_1\delta_1^{\,2} + N_2\delta_2^{\,2} + N_1d_1^{\,2} + N_2d_2^{\,2}}{N_1 + N_2}}$$

where δ_{12} = Combined standard deviation

δ_1 = standard deviation of the first group

δ_2 = standard deviation of the second group

$d_1 = \left|\bar{X}_1 - \bar{X}_{12}\right|$; $d_2 = \left|\bar{X}_2 - \bar{X}_{12}\right|$

NB: The above formula can be extended to find out the standard deviation of three or more groups. For example, combined standard deviation of three groups would be

$$\delta_{123} = \sqrt{\frac{N_1\delta_1^{\,2} + N_2\delta_2^{\,2} + N_3\delta_3 + N_1d_1^{\,2} + N_2d_2^{\,2} + N_3d_3}{N_1 + N_2 + N_3}}$$

Where $d_1 = \left|\bar{X}_1 - \bar{X}_{123}\right|$; $d_2 = \left|\bar{X}_2 - \bar{X}_{123}\right|$; $d_3 = \left|\bar{X}_3 - \bar{X}_{123}\right|$

Example

1. The number of workers employed, the mean wage per week and the standard deviation in each branch of a company are given below. Calculate the mean wages and standard deviation of all workers taken together for the factory.

Branch	No. of workers	Weekly mean wage	Standard deviation
A	50	1413	60
B	60	1420	70
C	90	1415	80

Advantages of the standard deviation

- It is rigidly defined and is based on all the observations of the series
- It is applied or used in other statistical techniques like correlation and regression analysis and sampling theory.

- It is possible to calculate the combined standard deviation of two or more groups.

Disadvantages of the standard deviation

- It cannot be used for comparing the dispersion of two or more series of observations given in different units.
- It gives more weight to extreme values.

Coefficient of Variation

The measures of dispersion which are expressed in terms of the original units of the observations are termed as absolute measures. Such measures are not suitable for comparing the variability of two distributions which are not expressed in the same units of measurements. Therefore it is better to use relative measure of dispersion obtained as ratios or percentages and are thus pure numbers independent of the unit of measurement.

Standard deviation is an absolute measure of dispersion and a relative measure based on the standard deviation is called the coefficient of variation. It is a pure number and suitable for comparing the variability, homogeneity or uniformity of two or more distributions. It is given as a percentage and calculated as

Coefficient of variation (CV) = $\frac{\sigma}{Mean}\times 100$

The lower the C.V the more consistent or stable the distribution is since the less the variability.

Example

Over a period of 3 months the daily number of components produced by two comparable machines was measured, giving the following statistics

Machine A: mean = 242.8; Standard deviation = 20.5

Machine B: mean = 281.3; Standard deviation = 23.0

Which machine has less variability in its performance?

4.5 Skewness and Kurtosis

- The term ‘skewness’ refers to lack of symmetry or departure from symmetry. When a distribution is not symmetrical it is called a skewed distribution.

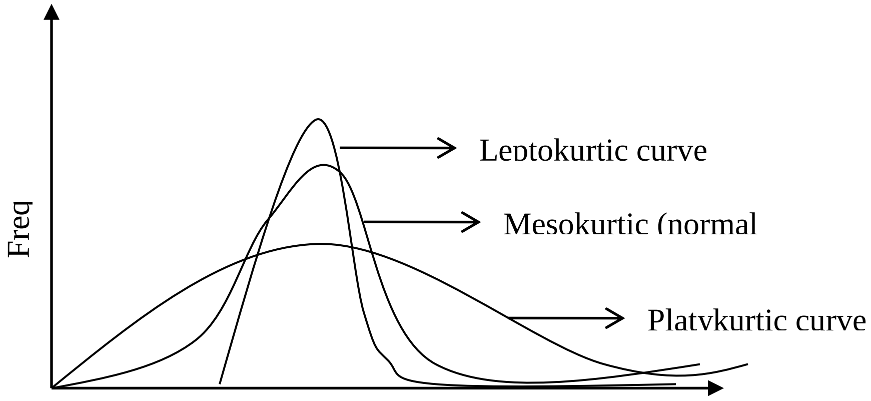
- In a symmetrical distribution the values of mean, median and mode are alike. If the value of mean is greater than the mode, skewness is said to be positive. If the value of mode is greater than mean, skewness is said to be negative.
- The Karl Pearson's coefficient of skewness is frequently used for measuring skewness and its calculated as

$$SK_p = \frac{Mean - Mode}{\delta}$$

But $Mean - Mode = 3(Mean - Median)$**.** Thus the formula for calculating the coefficient of skewness can be written as

$$SK_p = \frac{3(Mean - Median)}{\delta}$$

- Kurtosis refers to the degree of flatness or peakedness of a frequency curve. The degree of peakedness of a distribution is measured relative to the peakedness of the normal distribution.
- If a distribution is more peaked than the normal curve, it is called Leptokurtic; if it is more flat-topped than the normal curve, it is called platykurtic or flat-topped. The normal curve is itself known as Mesokurtic.



4.6 Activities

1. The following table indicates the marks obtained by students in a statistics test.

Marks	**Number of students**
0 – 20	5
20 – 40	7
40 – 60	-
60 – 80	8
80 – 100	7

The arithmetic mean for the class was 52.5 marks. You are required to determine the value of:

i) The missing frequency

ii) The median mark

iii) The modal mark

iv) The standard deviation

v) The coefficient of skewness

2. From the prices of the shares *X* and *Y* given below, state which share is more stable in value and which one would you invest on and why?

X:	55	54	52	53	56	58	52	50	51	49
Y:	108	107	105	105	106	107	104	103	104	101

3. An analysis of the monthly wages paid to workers of two firms A and B belonging to the same industry gives the following results:

	Firm A	**Firm B**
No. of wage earners	586	648
Average monthly wage	52.5	47.5
Standard deviation	10	11

Compute the combined standard deviation.