UNIT 8 STATISTICAL DERIVATIVES AND MEASURES OF CENTRAL TENDENCY

STRUCTURE

8.0 Objectives
8.1 Introduction
8.2 Statistical Derivatives
 8.2.1 Percentage
 8.2.2 Ratio
 8.2.3 Rate
8.3 Measures of Central Tendency
 8.3.1 Properties of an Ideal Measure of Central Tendency
 8.3.2 Mean and Weighted Mean
 8.3.3 Median
 8.3.4 Mode
 8.3.5 Choice of a Suitable Average
 8.3.6 Some Other Measures of Central Tendency
8.4 Let Us Sum Up
8.5 Key Words
8.6 Answers to Self Assessment Exercises
8.7 Terminal Questions/Exercises
8.8 Further Reading

8.0 OBJECTIVES

After studying this unit, you should be able to:

- explain the meaning and use of percentages, ratios and rates for data analysis,
- discuss the computational aspects involved in working out the statistical derivatives,
- describe the concept and significance of various measures of central tendency, and
- compute various measures of central tendency, such as arithmetic mean, weighted mean, median, mode, geometric mean, and hormonic mean.

8.1 INTRODUCTION

In Unit 6 we discussed the method of classifying and tabulating of data. Diagrammatic and graphic presentations are covered in the previous unit (Unit-7). They give some idea about the existing pattern of data. So far no big numerical computation was involved. Quantitative data has to be condensed in a meaningful manner, so that it can be easily understood and interpreted. One of the common methods for condensing the quantitative data is to compute statistical derivatives, such as Percentages, Ratios, Rates, etc. These are simple derivatives. Further, it is necessary to summarise and analyse the data. The first step in that direction is the computation of Central Tendency or Average, which gives a bird's-eye view of the entire data. In this Unit, we will discuss computation of statistical derivatives based on simple calculations. Further, numerical methods for summarizing and describing data – measures of Central Tendency – are discussed. The purpose is to identify one value, which can be obtained from the data, to represent the entire data set.

8.2 STATISTICAL DERIVATIVES

Statistical derivatives are the quantities obtained by simple computation from the given data. Though very easy to compute, they often give meaningful insight to the data. Here we discuss three often-used measures: percentage, ratio and rate. These measures point out an existing relationship among factors and thereby help in better interpretation.

8.2.1 Percentage

As we have noted earlier, the frequency distribution may be regarded as simple counting and checking as to how many cases are in each group or class. The relative frequency distribution gives the proportion of cases in individual classes. On multiplication by 100, the percentage frequencies are obtained. Converting to percentages has some advantages - it is now more easily understood and comparison becomes simpler because it standardizes data. Percentages are quite useful in other tables also, and are particularly important in case of bivariate tables. We show one application of percentages below. Let us try to understand the following illustration.

Illustration 1

The following table gives the total number of workers and their categories for all India and major states. Compute meaningful percentages.

Table: Total Workers and Their Categories-India and Major States : 2001
(*In thousands*)

Sl. No.	State/ India	Cultivators	Agricultural Labourers	Household Industry Workers	Other Workers	Total Workers
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1.	Jammu & Kashmir	1600	249	230	1611	3689
2.	Himachal Pradesh	1961	93	50	887	2991
3.	Punjab	2099	1499	307	5236	9142
4.	Haryana	3046	1276	207	3854	8383
5.	Rajasthan	13167	2529	651	7434	23781
6.	Uttar Pradesh	22173	13605	2886	15517	54180
7.	Bihar	8192	13528	1087	5273	28080
8.	Assam	3742	1290	329	4197	9557
9.	West Bengal	5613	7351	2153	14386	29503
10.	Orissa	4238	5001	689	4344	14273
11.	Madhya Pradesh	11059	7381	1010	6307	25756
12.	Gujarat	5613	4988	382	9386	20369
13.	Maharas-htra	12010	11291	1046	17706	42053
14.	Andhra Pradesh	7904	13819	1570	11573	34865
15.	Karnataka	6936	6209	936	9441	23522
16.	Kerala	740	1654	365	7532	10291
17.	Tamil Nadu	5114	8665	1459	12574	27812
	INDIA	127628	107448	16396	151040	402512

Solution: In the table above, the row total gives the total workers of a state/ all India and column total gives the aggregate values of different categories of workers and all workers. Thus, it is possible to compute meaningful percentages from both rows and columns. The row percentages are computed by dividing the figures in columns (3), (4), (5) and (6) by the figure in column (7) and multiplied by 100. The figures are presented in tabular form below. Percentage of cultivators in Jammu & Kashmir is obtained as $(1600 \div 3688) \times 100$ which equals 43.37. Similarly other figures are obtained.

Table: Percentage of Total Workers and Their Categories-India and Major States : 2001

Sl. No.	State/ India	Cultivators	Agricultural Labourers	Household Industry workers	Other wor-kers	Total Workers
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1.	Jammu & Kashmir	43.37	6.74	6.22	43.67	100.00
2.	Himachal Pradesh	65.55	3.10	1.68	29.67	100.00
3.	Punjab	22.96	16.40	3.36	57.28	100.00
4.	Haryana	36.34	15.22	2.47	45.97	100.00
5.	Rajasthan	55.36	10.64	2.74	31.26	100.00
6.	Uttar Pradesh	40.92	25.11	5.33	28.64	100.00
7.	Bihar	29.17	48.18	3.87	18.78	100.00
8.	Assam	39.15	13.50	3.44	43.91	100.00
9.	West	19.02	24.92	7.30	48.76	100.00
	Bengal					
10.	Orissa	29.70	35.04	4.82	30.44	100.00
11.	Madhya Pradesh	42.93	28.66	3.92	24.49	100.00
12.	Gujarat	27.56	24.49	1.87	46.08	100.00
13.	Maharas-htra	28.56	26.85	2.49	42.10	100.00
14.	Andhra Pradesh	22.67	39.63	4.51	33.19	100.00
15.	Karnataka	29.48	26.40	3.98	40.14	100.00
16.	Kerala	7.19	16.07	3.55	73.19	100.00
17.	Tamil Nadu	18.39	31.16	5.24	45.21	100.00
	INDIA	31.72	26.69	4.07	37.52	100.00

The figures above help in comparing the proportion of workers in different categories across the state and all India. One may read from the table that Kerala has the lowest percentage of cultivators and Bihar the highest percentage of agricultural labourers.

Self Assessment Exercise A

1) What is a Percentage?

....

....

....

....

2) From the data given in illustration 1, compute column percentages and interpret it. Why are the totals of these percentages not adding to 100?

The table below may be used for computation.

Table: State-wise Percentage Share of Total Workers and Categories of Workers in All India: 2001

Sl. No.	State/ India	Cultiva-tors	Agricultural Labourers	Household Industry workers	Other wor-kers	Total Workers
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1.	Jammu & Kashmir					
2.	Himachal Pradesh					
3.	Punjab					
4.	Haryana					
5.	Rajasthan					
6.	Uttar Pradesh					
7.	Bihar					
8.	Assam					
9.	West Bengal					
10.	Orissa					
11.	Madhya Pradesh					
12.	Gujarat					
13.	Maha-rashtra					
14.	Andhra Pradesh					
15.	Karnataka					
16.	Kerala					
17.	Tamil Nadu					
	INDIA	100.00	100.00	100.00	100.00	100.00

8.2.2 Ratio

Another descriptive measure that is commonly used with frequency distribution (it may be used elsewhere also) is the ratio. It expresses the relative value of frequencies in the same way as proportion or percentages but it does so by comparing any one group to either total number of cases or any other group. For instance, in table 6.3, Unit 6, the ratio of all labourers to their daily wages between Rs 30–35 is 70:14 or 5:1. Where ever possible, it is convenient to reduce the ratios in the form of n1: n2, the most preferred value of n2 being 1. Thus, representation in the form of ratio also reduces the size of the number which facilitates easy comparison and quick grasp. As the number of categories increases, the ratio is a better derivative for presentation as it will be easy and less confusing.

There are several types of ratios used in statistical work. Let us discuss them.

The Distribution Ratio: It is defined as the ratio of a part to a total which includes that part also. For example, in an University there are 600 girls out of 2,000 students. Than the distribution ratio of girls to the total number of students is 3:10. We can say 30% of the total students are girls in that University.

Interpret ratio: It is a ratio of a part in a total to another part in the same total. For example, sex ratio is usually expressed as number of females per 1,000 males (not against population).

Time ratio: This ratio is a measure which expresses the changes in a series of values arranged in a time sequence and is typically shown as percentage. Mainly, there are two types of time ratios :

i) **Those employing a fixed base period:** Under this method, for instance, if you are interested in studying the sales of a product in the current year, you would select a particular past year, say 1990 as the base year and compare the current year's production with the production of 1990.

ii) **Those employing a moving base:** For example, for computation of the current year's sales, last year's sales would be assumed as the base (for 1991, 1990 is the base. For 1992, 1991 is the base and so on …. .

Ratios are more often used in financial economics to indicate the financial status of an organization. Look at the following illustration:

Illustration 2

The following table gives the balance sheet of XYZ Company for the year 2002–03. Compute useful financial ratios.

Table: Balance Sheet of XYZ Company as on March 31, 2003

I	Sources of Funds	Amount (Rs. 000')
1	Shareholders' funds	520
1(a)	Share capital	130
1(b)	Reserve and surplus	390
2	Loan funds	280
2 (a)	Secured loans	170
2 (a) (i)	Due after one year	120
2 (a) (ii)	Due within one year	50
2 (b)	Unsecured loans	110
2 (b) (i)	Due after one year	50
2 (b) (ii)	Due within one year	60
	Total (520 + 280)	800

II	Application of Funds	Amount (in Rs. 000')
1	Net fixed asset	535
2	Investments	85
2 (a)	Long term investments	75
2 (b)	Current investments	10
3	Current assets, loans and advances	330
3 (a)	Inventories	160
3 (b)	Sundry debtors	80
3 (c)	Cash and bank balances	40
3 (d)	Loans and advances	50
	Less: Current liabilities and provisions	150
	Net current assets	180
	Total (535 + 85 + 180)	800

Solution: Three common ratios may be computed from the above balance sheet: current ratio, cash ratio, and debt-equity ratio. However, these ratios are discussed in detail in MCO-05 : Accounting for Managerial Decisions, under Unit-5 : Techniques of Financial Analysis.

$$\text{Current ratio} = \frac{\text{Current assets, loans, advances + current investments}}{\text{Current liabilities and provisions + short term debt}} = \frac{330 + 10}{150 + 50 + 60}$$

$$= 1.31$$

$$\text{Cash ratio} = \frac{\text{Cash and bank balances + Current investments}}{\text{Current liabilities and provisions + Short term debt}} = \frac{40 + 10}{150 + 50 + 60} = 0.19$$

$$\text{Debt} - \text{equity ratio} = \frac{\text{Debt}}{\text{Equity}} = \frac{\text{Loan fund}}{\text{Shareholders' funds}} = \frac{280}{520} = 0.54$$

8.2.3 Rate

The concept of ratio may be extended to the rate. The rate is also a comparison of two figures, but not of the same variable, and it is usually expressed in percentage. It is a measure of the number of times a value occurs in relation to the number of times the value could occur, i.e. number of actual occurrences divided by number of possible occurrences. Unemployment rate in a country is given by total number of unemployed person divided by total number of employable persons. It is clear now that a rate is different from a ratio. For example, we may say that in a town the ratio of the number of unemployed persons to that of all persons is 0.05: 1. The same message would be conveyed if we say that unemployment rate in the town is 0.05, or more commonly, 5 per cent. Sometimes rate is defined as number of units of a variable corresponding to a single unit of another variable; the two variables could be in different units. For example, seed rate refers to amount of seed required per unit area of land. The following table gives some examples of rates.

S.No.	Description	Computation	Rate
(1)	(2)	(3)	(4)
1	100 kms with 8 litres of petrol	100/8	12.5 km per litre
2	Rs. 18 for 12 banana	18/12	Rs. 1.50 per banana
3	Rs. 6000 for 5 days of consultancy	6000/5	Rs. 1200 per day consultancy

Self Assessment Exercise B

1) Name the different types of ratios used in statistical work.

....

....

....

2) What is a rate?

....

....

....

8.3 MEASURES OF CENTRAL TENDENCY

In Unit 6, we have studied in detail how to classify raw data into a small number of classes or groups and presented them in the form of tables. The next step would be to identify a single value that may be considered as the most representative value of the given data. This is the measure of central tendency, which represents an average character.

A measure of central tendency helps us to represent a set of huge data by a single value. To understand the economic condition of people of a particular country, we talk of average or per capita income. It also enables us to compare the situation in two different places or situations. For example, one may compare per capita power availability in two states to understand which one is better in terms of industrial climate.

To start with, we list the properties that could be defined by an ideal measure of central tendency. Some of the measures are discussed in detail later.

8.3.1 Properties of an Ideal Measure of Central Tendency

An ideal measure of central tendency should have the following properties:

- simple to compute and easy to interpret.
- based on all observations.
- should not be influenced much by a few observations.
- should be capable of further algebraic treatment.
- should be capable of being defined unambiguously.

Some of the important measures of central tendency which are most commonly used in business and industry are: Arithmetic Mean, Weighted Arithmetic Mean, Median, Mode, Geometric mean and Harmonic mean. Among them Median and Mode are the positional averages and the rest are termed as Methamatical Averages.

8.3.2 Mean and Weighted Mean

Most of the time, when we refer to the average of something, we are talking about the arithmetic mean. This is the most important measure of central tendencies which is commonly called **mean**.

Mean of ungrouped data: The mean or the arithmetic mean of a set of data is given by:

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{N}$$

This formula can be simplified as follows:

Arithmetic mean $(\overline{x}) = \frac{\sum x}{N}$ ← Sum of values of all observations. / ← Number of observations.

The Greek letter sigma, Σ, indicates "the sum of "

Illustration 3

Suppose that wages (in Rs) earned by a labourer for 7 days are 22, 25, 29, 58, 30, 24 and 23.The mean wage of the labourer is given by:

(22 + 25 + 29 + 58 + 30 + 24 + 23)/7 = Rs. 30.14

Mean of grouped data: We have seen how to obtain the result of mean from ungrouped data. In Unit-6, we have learnt the preparation of frequency distribution (grouped data). Let us consider what modifications are required for grouped data for calculation of mean.

When we have grouped data, either in the form of discrete or continuous, the expression for the mean would be :

$$(\overline{x}) = \frac{\sum fx}{N}$$

← Σ (f × x) / ← Sum of the frequency (Σf)

Let us consider an illustration to understand the application of the formula.

Illustration 4

The following discrete frequency distribution of wage data of a labourer for 35 days:

Wage	(in Rupees)	23	24	25	27	28	29	30	31	32	33	34
Frequency	(No of Days)	1	1	3	3	4	6	4	5	5	2	1

Now, to compute the mean wage, multiply each variable with its corresponding frequency (f × x) and obtain the total (Σfx).

Divide this total by number of observations (Σf or N). Practically, we compute the mean as follows:

$$\text{Mean} = \frac{(23\times1+24\times1+25\times3+27\times3+28\times4+29\times6+30\times4+31\times5+32\times5+33\times2+34\times1)}{(1+1+3+3+4+6+4+5+5+2+1)}$$

$$= \frac{102 \longrightarrow \Sigma fx}{35 \longrightarrow \Sigma f \text{ or } N} = 29.26$$

When a **frequency distribution** consists of data that are grouped by classes, it is known as continuous frequency distribution. In such a distribution each value of an observation falls somewhere in one of the classes. Unlike the raw data (ungrouped) or discrete data we do not know the seperate values of every observation. It is, therefore, to be noted that we can easily compute an estimate of the value of mean of continuous distribution but not the actual value of mean. On the other hand, we can say for ease of calculation, we cannot be very accurate.

To find the mean of continuous frequency distribution, we first calculate the midpoints of each class. Then we multiply each mid-point by the frequency of observations in that class, obtain sum of these products, and divide the sum by the total number of observations. The formula looks like this:

$$\bar{x} = \frac{\Sigma fx}{N}$$

where, Σfx = Sum value, which is obtained by multiplying the mid-points with its respective frequencies

N = Number of observations (Σf)

Let us consider the frequency distribution obtained in Unit-6 (table 6.3), as an illustration for study.

Illustration-5

The following table gives the daily wages for 70 labourers on a particular day.

Daily Wages (Rs) :	15-20	20-25	25-30	30-35	35-40	40-45	45-50
No of labourer :	2	23	19	14	5	4	3

Solution: For obtaining the estimated value of mean we have to follow the procedure as explained above. This is elaborated below.

Daily wages (Rs)	Mid-point (x)	No. of workers (f)	f.x
15–20	17.5	2	35.0
20–25	22.5	23	517.5
25–30	27.5	19	522.5
30–35	32.5	14	455.0
35–40	37.5	5	187.5
40–45	42.5	4	170.0
45–50	47.5	3	142.5
	–	N or Σf = 70	Σfx = 2030.0

$$\overline{X} = \frac{\sum fx}{N} = \frac{2030}{70} = \text{Rs. } 29$$

Hence, the mean daily wage is Rs. 29.

To simplify calculations, the following formula for mean may be more convenient to use. It is to be noted that it can be applied when the width of the classes are equal.

$$\overline{X} = A + \frac{\Sigma fd}{N} x\ i$$

where, 'A' is an assumed mean, $d = \frac{x - A}{i}$ and 'i' is the size of the equal class interval.

This formula makes the computations very simple and takes less time. This method eliminates the problem of large and inconvenient mid-points. To apply this formula, let us consider the data of the previous illustration-5. Try to understand the procedure, for obtaining the value of mean, shown below.

Assume A as 32.5

Class Interval	**Mid-point (X)**	**(X-32.5)/5 = d**	**Frequency (f)**	**fd**
15-20	17.5	– 3	2	– 6
20-25	22.5	– 2	23	– 46
25-30	27.5	– 1	19	– 19
30-35	32.5	0	14	0
35-40	37.5	1	5	5
40-45	42.5	2	4	8
45-50	47.5	3	3	9
	–	–	N = 70	Σfd = – 49

$$\overline{X} = A + \frac{\sum fd}{N} \times i$$

$$= 32.5 + \frac{-49}{70} \times 5 = 29$$

Hence mean daily wage is Rs. 29, as obtained earlier.

The important **property of arithmetic mean** is that the means of several sets of data may be combined into a single mean for the combined sets of data. The combined mean may be defined as:

$$\overline{X}_{12...n} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2 ..... + N_n\overline{X}_n}{N_1 + N_2 ..... + N_n}$$

If we have to combine means of four sets of data, then the above formula can be generalized as:

$$\overline{X}_{1234} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2 + N_3\overline{X}_3 + N_4\overline{X}_4}{N_1 + N_2 + N_3 + N_4}$$

Advantages and disadvantages of mean

The concept of mean is familiar to most people and easily understood. It is due to the fact that it possesses almost all the properties of a good measure of central tendency. However, the mean has disadvantages of which we must be aware. First, the value of mean may be distorted by the presence of extreme values in a given data set and in case of U-shaped distribution this measure is not likely to serve a useful purpose. Second problem with the mean is that we are unable to compute mean for open-ended classes, since it is difficult to assign a mid-point to the open-ended classes. Third, it cannot be used for qualitative variables.

Weighted Mean

The arithmetic mean, as discussed above, gives equal importance (weight) to all the observations. But in some cases, all observations do not have the same weightage. In such a case, we must compute weighted mean. The term 'weight', in statistical sense, stands for the relative importance of the different variables. It can be defined as:

$$\bar{x}_W = \frac{\sum Wx}{\sum W}$$

where, $\bar{x}_w$ is the weighted mean, 'w' are the weights assigned to the variables (x).

Weighted mean is extensively used in Index numbers, it will be discussed in detail in Unit 12 : Index Numbers, of this course. For example, to compute the cost of living index, we need the price index of different items and their weightages (percentage of consumption). The important issue that arises is the selection of weightages. If actual weightages are not available then estimated or arbitrary weightages may be used. This is better than no weightages at all. However, keeping the phenomena in mind, the weightages are to be assigned logically. To understand this concept, let us take an illustration.

Illustration 6

Given below are Price index numbers and weightages for different group of items of consumption for an average industrial worker. Compute the cost of living index.

Group Item	Group Price Index	Weight
Food	150	55
Clothing	186	15
House rent	125	17
Fuel and Light	137	8
Others	184	5

Solution: The cost of living index is obtained by taking the weighted average as explained in the table below:

Group Item	Group Price Index(Pi)	Weight (Wi)	Wi. Pi
Food	150	55	8250
Clothing	186	15	2790
House rent	125	17	2125
Fuel and Light	137	8	1096
Others	184	5	920
	–	ΣW = 100	ΣWx = 15181

Therefore, the cost of living index is $\overline{X}_W = \frac{\sum Wx}{\sum W} = \frac{15181}{100} = 151.81$

Self Assessment Exercise C

1) A student's marks in a Computer course are 69, 75 and 80 respectively in the papers on Theory, Practical and Project Work.

   What are the mean marks if the weights are 1, 2 and 4 respectively?

   What would be the mean marks if all the papers have equal importance? Use the following table

Table: Computation of Weighted Mean Marks

Paper	**Marks Percentage (X)**	**Weight (W)**	**W. X**
Theory			
Practical			
Project Work			

So, weighted mean is =

2) The following table gives frequency distribution of monthly sales (in Rupees thousands) of 125 firms.

Table: Monthly Sales of 125 Firms

Monthly Sales (in thousands)	Number of Firms
0–150	15
150–300	22
300–450	64
450–600	11
600–750	9
750–900	4
All	125

Compute mean monthly sales of the firms and interpret the data.

Since the class width is 150 for all the classes, the method of assumed mean is useful. The following table may be helpful.

Table: Computation of Average Monthly Sales of 125 Firms

Monthly Sales (in thousands)	Midpoint (X)	(X–A)/150 (d)	Number of Firms (f)	f.d

So, the average monthly sales =

3) The mean wage of 200 male workers in a factory was Rs 150 per day, and the mean wage of 100 female and 50 children were Rs. 90 and Rs. 35 respectively, in the same factory. What would be the combined mean of the workers. Comment on the result.

....

....

....

....

8.3.3 Median

The median is another measure of central tendency. The median represents the middle value of the data that measures the central item in the data. Half of the items lie above the median, and the other half lie below it.

Median of Ungrouped Data: To find the median from ungrouped data, first array the data either in ascending order or in descending order. If the number of observations (N) contains an odd number then the median is the middle value. If it is even, then the median is the mean of the middle two values.

In formal language, the median is $\left(\frac{N+1}{2}\right)^{th}$ item in a data array, where N is the number of items. Let us consider the earlier illustration 3 to locate the median value in two different sets of data.

Illustration-7

On arranging the daily wage data of the labourers (as given in illustration 3) in ascending order, we get

Rs. 22, 23, 24, 25, 29, 30, 58.

Number of observations is an odd number (seven). According to equation $\left(\frac{N+1}{2}\right)^{th}$ Item, the middle (i.e. the fourth) number is the median. Here the median wage of labourer is Rs. 25.

You may notice that unlike the mean we calculated earlier, the median we calculated above was not distorted by the presence of the last value i.e., Rs. 58. This value could have been even Rs. 99, the median would have been the same.

Had there been one more observation, say, Rs. 6, the order would have been as below:

Rs. 6, 22, 23, 24, 25, 29, 30, 58

There are eight observations, and the median is given by the mean of the fourth and the fifth observations (i.e., $\frac{8+1}{2}$th item = 4.5th item. So, median wage = (24 + 25)/2 = Rs. 24.5.

Median of Grouped Data: Now, let us calculate the median from grouped data. When the data is in the form of **discrete series**, the median can be computed by examining the cumulative frequency distribution, as is shown below.

Illustration-8

To compute median wage from the data given in Illustration 4, we add one more row of cumulative frequency (the formation of cumulative frequency, we have discussed in Unit 6 of this course : Processing of Data.

Wage (In Rupees)	23	24	25	27	28	29	30	31	32	33	34
Frequency (No. of Days)	1	1	3	3	4	6	4	5	5	2	1
Cumulative Frequency	1	2	5	8	12	18	22	27	32	34	35

According to the formula, $\left(\frac{N+1}{2}\right)^{th}$ item, the number of observations is 35.

Therefore, item $\frac{35+1}{2}$th is the 18th item. Hence the 18th observation will be the median. By inspection it is clear that Median wage is Rs. 29.

This procedure is to be slightly modified for computation of median from **class interval data**. The median is taken as the value of the variable corresponding to the (N/ 2)th observation. The class or group containing median should be identified first and the median is computed under the assumption that all the observations in that class are equally spaced. Symbolically, the expression for median is given by:

$$\text{Median} = L + \frac{N/2 - c.f}{f} \times i$$

where, 'L' is the lower limit of the median class, 'N' is the number of observations, 'f' is frequency of the median class, 'cf' is the cumulative frequency of the class next lower to the median class and 'i' is the width of the median class. Let us consider the data given in earlier illustration 5 to study the median.

Illustration 9: Compute median wage for the data given in Illustration 5.

Solution: The approach is quite similar to the previous example. As indicated, it will be implicitly assumed that the wages of the labourers in the group-containing median are equally spaced.

Class Interval (wages in Rs.)	Frequency (f)	Cumulative frequency (cf)
15-20	2	2
20-25	23	25
25-30	19	44
30-35	14	58
35-40	5	63
40-45	4	67
45-50	3	70
	N = 70	–

Here, the number of observations is 70. So the median corresponds to the 35th value of the variable $\left(\frac{70}{2}\right)^{th}$ item. This item lies in 44 (35th observation) cumulative frequency. Hence, It is clear Column (3) that median is in the third class interval, i.e. Rs. 25 to Rs. 30. So we have to locate the position of the 35th observation in the class 25-30.

Here, $\frac{N}{2} = 35$, L = 25, cf = 25, f = 19 and i = 30 – 25 = 5.

Thus median is : $L + \frac{N/2 - cf}{f} \times i = 25 + \frac{35-25}{19} \times 5 = Rs\,27.63.$

It is to be noted that the median value may also be located with the help of graph by drawing ogives or a less than cumulative frequency curve. This method was discussed in detail in Unit 7 : Diagrammatic and Graphic Presentation, of this block.

Advantages and disadvantages of median

The biggest advantage of median is that extreme observations do not affect it. For computation of the median, it is not necessary to know all the observations and this property comes in handy when there are open-ended classifications of data. This is also suitable for qualitative variables, which can be ordered or arranged in ascending or descending order (ordinal variables).

However, it requires data to be arranged before computation. It is not amenable to arithmetic and algebraic manipulations. For example, if M_1 and M_2 are medians of two different sets of data, we cannot get the median of the combined data set from M_1 and M_2.

Some Additional Points: Median divides the distribution in two equal parts. When a distribution is divided into four equal parts they are called **quartiles**. Similarly, there are **deciles** (divided into ten equal parts), **percentiles** (divided into hundred equal parts) etc. The general term for all of them is fractile. In Unit 9 of this block, we will learn more about quartiles.

Self Assessment Exercise D

1) Refer to the data in Self Assessment Exercise C, No. 1. Obtain median monthly sales of the firms.

The following table may be helpful.

Table: Computation of Median Sales of 125 Firms

Monthly Sales (in thousands)	Number of Firms (f)	Cumulative frequency

....

....

8.3.4 Mode

Mode is also a measure of central tendency. This measure is different from the arithmetic mean, to some extent like the median because it is not really calculated by the normal process of arithmetic. The mode, of the data, is the value that appears the maximum number of times. In an **ungrouped data**, for example, the foot size (in inches) of ten persons are as follows: 5, 8, 6, 9, 11, 10, 9, 8, 10, 9. Here the number 9 appeares thrice. Therefore, mode size of foot is 9 inches. **In grouped data** the method of calculating mode is different between discrete distribution and continous distribution.

In discrete data, for example consider the earlier illustration 6, the modal wage is Rs. 29 as is the wage for maximum number of days, i.e. six days. For continuous data, usually we refer to modal class or group as the class with the maximum frequency (as per observation approach). Therefore, the mode from continuous distribution may be computed using the expression:

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

where, L = lower limit of the modal class, i = width of the modal class, Δ_1 = excess of frequency of the model class (pi) over the frequency of the preceding class (f_0),
Δ_2 = excess of frequency of the model class (f_1) over the frequency of the succeeding class (f_2). The letter Δ is read as delta.

Noting that, $\Delta_1 = f_1 - f_0$ and $\Delta_2 = f_1 - f_2$

It is to be noted that while using the formula for mode, you must arrange the class intervals uniformly throughout, otherwise you will get misleading results. To illustrate the computation of mode, let us consider the grouped data of earlier illustration 7.

Illustration 10

Compute mode from the following data.

Daily wages (Rs.)	No. of workers (f)	Daily wages (Rs)	No. of workers (f)
15-20	12	35-40	5
20-25	23	40-45	4
25-30	19	45-50	3
30-35	14		

Solution: Since the maximum frequency 23 is in the class 20-25. Therefore, based on observation method, the class 20-25 is the modal class. Applying the

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

formula, we get $\Delta_1 = f_1 - f_0 ; \Delta_2 = f_1 - f_2$

The related values are as follows:

L = 20; f_0 = 12; f_1= 23; f_2 = 19; and i = 5 $\therefore \Delta_1 = 11 \,\&\, \Delta_2 = 4$

$$\therefore \text{ Mode} = 20 + \frac{11}{11+4} \times 5 = 20 + \frac{11}{15} \times 5$$

$$= 20 + 3.67 = \text{Rs. } 23.67$$

Hence the modal daily wage is Rs. 23.67

In a continuous frequency distribution, the value of mode can also be located graphically. We have already discussed the procedure for locating the mode graphically in Unit-7 of this block.

Advantages and Disadvantages of Mode

Extreme values of observations do not affect the mode and its value can be determined in open-ended classes. This measure is also suitable for any qualitative variables (both nominal and ordinal variables).

It may not be unique all the time. There may be more than one mode or no mode (no value that occurs more than once) at all. In such a case it is difficult to interpret and compare the distributions. It is not amenable to arithmetic and algebraic manipulations. For example, we cannot get the mode of the combined data set from the modes of the constituent data sets.

Self Assessment Exercise E

Refer to the data in Self Assessment Exercise C No. 1. Obtain mode of monthly sales of the firms.

L = , f_0 = , f_1 = , f_2 = and i = .

Hence, the mode is given by :

....

....

....

Comparing the Mean, Median, and Mode

For a moderately skewed distribution, it has been empirically observed that the difference between Mean and Mode is approximately three times the difference between Mean and Median. This was illustrated in the Fig. 8.1 (b) and (c). The expression is:

Mean – Mode = 3(Mean – Median)

Alternately, Mode = 3(Median) –2(Mean)

Sometimes this expression is used to calculate value of one measure when the value of the other two measures are known.

8.3.5 Choice of a Suitable Average

We have already discussed advantages and disadvantages of three different types of averages: Mean, Median, and Mode. Here, we discuss their appropriateness in terms of the following three factors: (1) the level of measurement of data (2) the shape of the distribution and (3) the stability of the measure of the average.

Levels of measurement: There are four levels of measurement of data: nominal, ordinal, ratio and interval. At **nominal level**, the observations can be just distinguished or differentiated but cannot be arranged in any order. Examples may be colour of cars, types of blood groups, brands of a consumer goods etc. At **ordinal level**, the observations can be arranged in ascending or descending order, but no arithmetic operations are possible. While describing the existing business climate, the respondents may tell - very good, good, medium, bad and very bad. This could be an example of ordinal data. At **interval level**, it is assumed that a given interval on the scale measures the same amount of difference, irrespective of where the interval appears. There is a zero but it is arbitrary and is not of much significance. For example, the temperature difference between 50^0C and 60^0C is the same, as the temperature difference between 10^0C and 20^0C but a temperature of 0^0C does not mean absence of heat. Variables like height, weight are examples of **ratio levels** of measurement. Here, a value which is twice as large as another value corresponds to twice the value of the variable and it has an absolute zero. We say that a 10-metre tower is twice as tall as a 5-metre tower, but we never mean that a temperature of 40^0C is twice as hot as a temperature of 20^0C.

From the above discussion, it is clear that for nominal data only mode can be used, for ordinal data both mode and median can be used whereas for ratio and interval levels of data all three measures can be calculated.

Shape of the distribution: If the distribution of data is symmetric with only one peak, mean, median, and mode are the same. Even in case of two modes, mean and median will be the same. For asymmetric distribution, all these are different. For positively skewed distribution, mode is the smallest and median lies between mode and mean whereas for negatively skewed distribution the pattern is just the opposite. (It is discussed elaborately in Unit 9, Section 9.5.) Thus, in either of the cases, median appears to be a better measure of central tendency. Figure 8.1 shows three different shapes of a distribution.

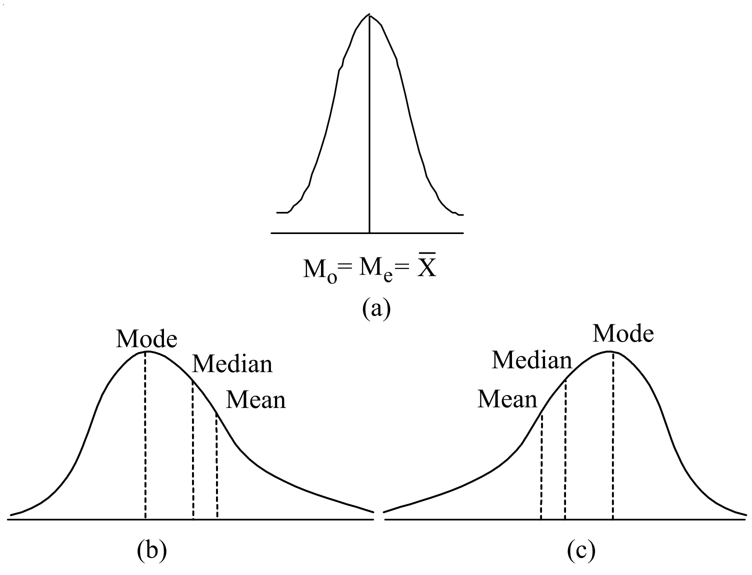

Figure 8.1

Stability: Quite often a researcher studies a sample to infer about the entire population. Mean is generally more stable than median or mode. If we calculate means, medians and modes of different samples from a population, the means will generally be more in agreement then the medians or the modes. Thus, mean is a more reliable measure of central tendency. Normally, the choice of a suitable measure of central tendency depends on the common practice in a particular industry. According to its requirement, each case must be judged independently. For example, the Mean sales of different products may be useful for many business decisions. The median price of a product is more useful to middle class families buying a new product. The mode may be a more useful measure for the garment industry to know the modal height of the population to decide the quantity of garments to be produced for different sizes.

Hence, the choice of the measure of central tendency depends on (1) type of data (2) shape of the distribution (3) purpose of the study. Whenever possible, all the three measures can be computed. This will indicate the type of distribution.

8.3.6 Some Other Measures of Central Tendency

Sometimes two other measures of central tendency, geometric mean and harmonic mean are also used. They are briefly discussed here.

Geometric Mean (GM): Geometric mean is defined as the Nth root of the product of all the N observations. It may be expressed as:

$\text{G.M.} = \sqrt[n]{\text{Product of all n values}}$. Thus, the geometric mean of four numbers 2, 5, 8 and 10 is given by $\sqrt[4]{(2\times5\times8\times10} = \sqrt[4]{(800)} = 5.3183$. If one observation is zero, the geometric mean becomes zero and hence inappropriate. If some values are negative, sometimes the geometric mean may be computed but may be meaningless. Geometric mean is appropriate for the variables that reproduce themselves. Suppose, population of a country in years 1990 and 2000 are respectively 100 and 121 million. The average population in the decade is $= \sqrt[2]{(100\times121)}$ million or 110 million. Probably the most frequent use of geometric mean is to know the average rate of change. These could be average percent change of population, compound interest, growth rate etc.

Harmonic Mean (HM): Harmonic mean is defined as the reciprocal of the arithmetic mean of the reciprocal of the observations. In other words, it may be defined as the ratio of Number of observations and sum of reciprocal of the values. It may be expressed as: $\text{HM} = N / \left(\frac{1}{x_1} + \frac{1}{x_2} + \ldots\ldots \frac{1}{x_n}\right)$, in short $N / \Sigma\left(\frac{1}{x}\right)$ For example, the harmonic mean of 4 and 6 is 2 / (1/4 + 1/6) = 2 / (5/12) = 2/ (5/12) = 21 (0.4166) = 4.8. Suppose a car moves half the distance at the speed of 60 km/hr and the other half at the speed of 80 km/hr. Then the average speed of the car is 68.57 km/hr, which is the harmonic mean of 60 and 80. Harmonic mean is useful in averaging rates.

For any set of data wherever computation is possible, the following inequality holds

$$\bar{x} > \text{GM} > \text{HM}$$

Illustration-11

To compute arithmetic, geometric and harmonic means of 4,5,10 and 11 and verify the above relationship.

Arithmetic Mean, A = (4 + 5 +10 +11)/4 = 30 / 4 = 7.50

Geometric Mean, G = $\sqrt[4]{(4\times5\times10\times11} = \sqrt[4]{(2200)} = 6.85$

Harmonic Mean, H = 4/(1/4 +1/5+1/10+1/11) = 4 /0.64 = 6.25

So, the relationship discussed above is verified.

It is also possible to compute weighted geometric and harmonic means.

8.4 LET US SUM UP

In order to draw meaningful and useful conclusions from the data the collected data must be analysed with the help of statistical derivatives like percentage, ratio and rates. They also give meaningful insight with very little computation. A ratio expresses the relationship between the magnitude of more than one quantity. It is generally stated as A : B : C. Proportion is the ratio of any one category to the total of all the categories. It is a better derivative to use when the number of categories increases. A rate is usually expressed as per 100 per 1,000 etc. A measure of central tendency gives one representative value, around which the data set is clustered. Three widely used measures are discussed in detail. Mode is the simplest of all but at times it is not defined. Median divides the observations into two equal parts and is particularly suitable in open-ended data. Arithmetic mean is calculated based on all the observations but gets affected by extreme values. For qualitative data, however mean cannot be computed. Mean, median and mode show the type of distribution of data.Measures of central tendency are also called measures of location.

8.5 KEY WORDS

Arithmetic Mean : This equals the sum of all the values divided by the number of observations.

Bimodal Distribution : In a distribution, when two values occur more frequently in equal number.

Geometric Mean : It refers to Nth root of the product of all the N observations.

Harmonic Mean : This is the reciprocal of the arithmetic mean of the reciprocals of the given values.

Mean : Usually refers to Arithmetic Mean or Average.

Median : The middlemost observation in a data set when arranged in order.

Mode : The most frequent value occurring in a data set. It is represented by the highest point in the distribution curve of a data set.

Percentage : It gives the magnitude of the numerator when denominator of a ratio becomes hundred.

Rate : Amount of one variable per unit amount of some other variable.

Ratio : Relative value of one value with respect to another value.

Weighted Mean : An average in which each observation value is weighted by some index of its importance.

8.6 ANSWERS TO SELF ASSESSMENT EXERCISES

A: 2. With reference to the original table, the all India figures are the totals of all the state figures. Thus a column percentage gives the share of a state from among all the states of India in respect of the category of workers. The column percentages are given below.

Table: State-wise Percentage Share of Total Workers and Categories of Workers in All India: 2001

Sl. No.	State/ India	Cultivators	Agricultural Labourers workers	Household Industry	Other Wor-kers	Total
1	Jammu & Kashmir	1.25	0.23	1.40	1.07	0.92
2	Himachal Pradesh	1.54	0.09	0.31	0.59	0.74
3	Punjab	1.64	1.40	1.87	3.47	2.27
4	Haryana	2.39	1.19	1.26	2.55	2.08
5	Rajasthan	10.32	2.35	3.97	4.92	5.91
6	Uttar Pradesh	17.37	12.66	17.60	10.27	13.46
7	Bihar	6.42	12.59	6.63	3.49	6.98
8	Assam	2.93	1.20	2.00	2.78	2.37
9	West Bengal	4.40	6.84	13.13	9.52	7.33
10	Orissa	3.32	4.65	4.20	2.88	3.55
11	Madhya Pradesh	8.66	6.87	6.16	4.18	6.40
12	Gujarat	4.40	4.64	2.33	6.21	5.06
13	Maharashtra	9.41	10.51	6.38	11.72	10.45
14	Andhra Pradesh	6.19	12.86	9.57	7.66	8.66
15	Karnataka	5.43	5.78	5.71	6.25	5.84
16	Kerala	0.58	1.54	2.22	4.99	2.56
17	Tamil Nadu	4.01	8.06	8.90	8.32	6.91
	INDIA	100.00	100.00	100.00	100.00	100.00

The interpretation is obvious - Out of all the workers in all India, 13.46 percent are in Uttar Pradesh and 10.45 percent in Maharashtra. Andhra Pradesh has the highest number of Agricultural Labourers (12.86%) followed by Uttar Pradeh (12.66%) and Bihar (12.59%). The lowest number of Household Industry workers is in Himachal Pradesh, etc.

C: 1) The weighted mean is = 539/7 = 77

If all the papers have equal importance, i.e. equal weightage, then the simple mean = 224/3 =74.67.

2) Since the class width is 150 for all the classes, the method of assumed mean is useful. On observation, assumed mean is taken as 375.

$\bar{x} = A + \frac{\Sigma fd}{N} xi$; Mean sales of 125 firms is Rs. 361.8 thousands.

3) $\bar{x}_{123} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2 + N_3\bar{x}_3}{N_1 + N_2 + N_3}$

$\bar{x}_{123} = 116.43$

D: Median $= L + \frac{N/2 - c.f}{f} \times c$

$M_e = 359.76$

E: $\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$; Modal sales value is Rs. 363.64 thousands.

8.7 TERMINAL QUESTIONS/EXERCISES

1) Explain the concept of central tendency with the help of an example. What purpose does it serve?
2) "A representative value of a data set is a number indicating the central value of that data". To what extent is it true for Mean, Median, and Mode? Explain with illustrations.
3) Discuss the merits and limitations of various measures of central tendency.
4) The following table gives workers of India (in thousands) as per 2001 census. Compute suitable percentages and interpret them.

Table: Total Workers and Their Categories-India : 2001 ***(In thousands)***

S. No.	Total Rural/ Urban	Persons Males/ Females	Cultivators	Agricultural Labourers	Household Industry workers	Other Workers	Total Workers
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	Rural	Persons	124682	103122	11710	71142	310655
2		Males	84047	54749	5642	54762	199200
3		Females	40635	48373	6067	16380	111456
4	Urban	Persons	2946	4326	4686	79899	91857
5		Males	2282	2605	2670	68707	76264
6		Females	664	1721	2016	11191	15593
7	Total	Persons	127628	107448	16396	151040	402512
8		Males	86328	57354	8312	123469	275464
9		Females	41300	50093	8084	27571	127048

Source: Census of India, 2001.

5) The monthly salaries (in Rupees) of 11 staff members of an office are: 2000, 2500, 2100, 2400, 10000, 2100, 2300, 2450, 2600, 2550 and 2700.

Find mean, median and mode of the monthly salaries.

Which one among the above do you consider the best measure of central tendency for the above data set and why?

6) Consider the data set given in problem 2 above.

Find mean deviation of the data set from (i) median (ii) 2400 and (iii) 2500.

Find mean squared deviation of the data set from (i) mean (ii) 3000 and (iii) 3100.

7) Mean examination marks in Mathematics in three sections are 68, 75 and 72, the number of students being 32, 43 and 45 respectively in these sections. Find the mean examination marks in Mathematics for all the three sections taken together.

8) The followings are the volume of sales (in Rupees) achieved in a month by 25 marketing trainees of a firm:

1220	1280	1700	1400	400	350	1200	1550	1300	1400
1450	300	1800	200	1150	1225	1300	1100	450	1200
1800	475	1200	600	1200					

The firm has decided to give the trainees some performance bonus as per the following rule - Rs. 100 if the volume of sales is below Rs. 500; Rs. 250 if the volume of sales is between Rs. 500 and Rs.1000; Rs.400 if the volume of sales is between Rs. 1000 and Rs, 1500 and Rs.600 if the volume of sales is above Rs. 1500.

Find the average value of performance bonus of the trainees.

9) In an urban cooperative bank, the minimum deposit in a savings bank is Rs. 500. The deposit balance at the end of a working day is given in the table below :

Table: Average Deposit Balance in ABC Urban Cooperative Bank

S No	Deposit Balance	Number of Deposits
1	Less than Rs. 10000	982
2	Less than Rs. 9000	959
3	Less than Rs. 8000	874
4	Less than Rs. 7000	773
5	Less than Rs. 6000	621
6	Less than Rs. 5000	395
7	Less than Rs. 4000	295
8	Less than Rs. 3000	145
9	Less than Rs. 2000	25
10	Less than Rs. 1000	10

Calculate mean, median and mode from the above data.

10) Refer to the table given in the previous problem. Compute (a) median and (b) mode by graphical approach.

11) Refer to the problem 8. Compute the approximate value of mode using the relationship: Mean - Mode = 3 (Mean - Median), and compare with the computed value obtained earlier.

Note: These questions/exercises will help you to understand the unit better. Try to write answers for them. But do not submit your answers to the university for assessment. These are for your practice only.

8.8 FURTHER READING

The following text books may be used for more indepth study on the topics dealt with in this unit.

Gupta, S P and M P Gupta, 1988. *Business Statistics*, S Chand, New Delhi.

Hooda, R.P., 2001. Statistics for Business and Economics, Macmillan India Limited, New Delhi.

Levin, R I and D S Rubin, 1998. *Statistics for Management*, Prentice Hall India, New Delhi.

Spiegel, M R, 1992. *Statistics*, Schaum's Outline Series, McGraw Hill, Singapore.