UNIT 7 MEASURES OF CENTRAL TENDENCY

Objectives

After going through this unit, you will learn:

- the concept and significance of measures of central tendency
- to compute various measures of central tendency, such as arithmetic mean, weighted arithmetic mean, median, mode, geometric mean and harmonic mean
- to compute several quantiles such as quartiles, deciles and percentiles
- the relationship among various averages.

Structure

7.1 INTRODUCTION

With this unit, we begin our formal discussion of the statistical methods for summarising and describing numerical methods for summarising and describing numerical data. The objective here is to find one representative value which can-be used to locate and summarise the entire set of varying values. This one value can be used to make many decisions concerning the entire set. We can define measures of central tendency (or location) to find some central value around which the data tend to cluster.

7.2 SIGNIFICANCE OF MEASURES OF CENTRAI TENDENCY

Measures of central tendency i.e. condensing the mass of data in one single value, enable us to get an idea of the entire data. For example, it is impossible to remember the individual incomes of millions of earning people of India. But if the average income is obtained, we get one single value that represents the entire population.

Measures of central tendency also enable us to compare two or more sets of data to facilitate comparison. For example, the average sales figures of April may be compared with the sales figures of previous months.

7.3 PROPERTIES OF A GOOD MEASURE OF CENTRAL TENDENCY

A good measure of central tendency should possess, as far as possible, the following properties,

i) It should he easy to understand.

ii) It should he simple to compute.

iii) It should be based on all observations.

iv) It should be uniquely defined.

v) It should be capable of further algebraic treatment.

vi) It should not be unduly affected by extreme values.

Following are some of the important measures of central tendency which are commonly used in business and industry.

Arithmetic Mean

Weighted Arithmetic Mean

Median

Quantiles

Mode

Geometric Mean

Harmonic Mean

7.4 ARITHMETIC MEAN

The arithmetic mean (or mean or average) is the most commonly used and readily understood measure of central tendency. In statistics, the term average refers to any of the measures of central tendency. The arithmetic mean is defined as being equal to the sum of the numerical values of each and every observation divided by the total number of observations. Symbolically, it can be represented as:

$$\bar{x} = \frac{\sum x}{N}$$

where $\sum x$ indicates the sum of the values of all the observations, and N is the total number of observations. For example, let us consider the monthly salary (Rs.) of 10 employees of a firm

2500, 2700, 2400, 2300, 2550, 2650, 2750, 2450, 2600, 2400

If we compute the arithmetic mean, then

$$\bar{x} = \frac{2500+2700+2400+2300+2550+2650+2750+2450+2600+2400}{10}$$

$$= \frac{25300}{10} = \text{Rs. } 2530.$$

Therefore, the average monthly salary is Rs. 2530.

We have seen how to compute the arithmetic mean for ungrouped data. Now let us consider what modifications are necessary for grouped data. When the observations are classified into a frequency distribution, the midpoint of the class interval would be treated as the representative average value of that class. Therefore, for grouped data; the arithmetic mean is defined as

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

Where X is midpoint of various classes, f is the frequency for corresponding class and N is the total frequency, i.e. $N = \sum f$.

This method is illustrated for the following data which relate to the monthly sales of 200 firms.

Monthly Sales (Rs. Thousand)	No. of Firms	Monthly Sales (Rs. Thousand)	No. of Firms
300-350	5	550-600	25
350-400	14	600-650	22
400-450	23	650-700	7
450-500	50	700-750	2
500-550	52		

For computation of arithmetic mean, we need the following table:

Monthly Sales (Rs. Thousand)	Mid point X	No. of firms f	fX
300-350	325	5	1625
350-400	375	14	5250
400-450	425	23	9775
450-500	475	50	23750
500-550	525	52	27300
550-600	575	25	14375
600-650	625	22	13750
650-700	675	7	4725
700-750	725	2	1450

$$\bar{x} = \frac{\sum fx}{N} = \frac{102000}{200} = 510$$

Hence the average monthly sales are Rs. 510.

To simplify calculations, the following formula for arithmetic mean may be more convenient to use.

$$\bar{x} = A + \frac{\sum fd}{N} \times i$$

where A is an arbitrary point, $d = \frac{X-A}{i}$, and i = size of the equal class interval.

REMARK: A justification of this formula is as follows. When $d = \frac{X-A}{i}$, then X = A + i d Multiplying throughout by F, taking summation on both sides and dividing by N, we get

$$\bar{x} = A + \frac{\sum fd}{N} \times i$$

This formula makes the computations very simple and takes less time. To apply this formula, let us consider the same example discussed earlier and shown again in the following table.

Monthly Sales (Rs. Thousand)	Mid point	No. of Firms f	(X-525)/50 =d	fd
300-350	325	5	-4	-20
350-400	375	14	-3	-42
400-450	425	23	-2	-46
450-500	*475*	*50*	-1	-50
500-550	525	52	0	0
550-600	575	25	+1	+25
600-650	625	22	+2	+44
650-700	675	7	+3	+21
700-750	725	2	+4	+8
		N=200		$\sum fd = 60$

$$\bar{x} = A + \frac{\sum fd}{N} \times i = 525 - \frac{60}{200} \times 50$$

$$= 525 - 15 = 510 \text{ or Rs. } 510$$

It may be observed that this formula is much faster than the previous one and the value of arithmetic mean remains the same.

7.5 MATHEMATICAL PROPERTIES OF ARITHMETIC MEAN

Because the arithmetic is defined operationally, it has several useful mathematical properties. Some of these are:

1) The sum of the deviations of the observations from the arithmetic mean is always zero. Symbolically, it is:

$$\sum (x - \bar{x}) = 0$$

It is because of this property that the mean is characterised as a point of balance, i.e, the sum of the positive deviations from mean is equal to the sum of the negative deviations from mean.

2) The sum of the squared deviations of the observations from the mean is minimum, i.e., the total of the squares of the deviations from any other value than the mean value will be greater than the total sum of squares of the deviations from mean. Symbolically,

$$\sum (x - \bar{x})^2 \text{ is a minimum.}$$

3) The arithmetic means of several sets of data may be combined into a single arithmetic mean for the combined sets of data. For two sets of data, the combined arithmetic mean may be defined as

$$\bar{x}_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$$

Where $\bar{X}_{12}$ = combined mean of two sets of data.

$\bar{X}_1$ = arithmetic mean of the first set of data.

$\bar{X}_2$ = arithmetic mean of the second set of data.

N_1 = number of observations in the first set of data.

N_2 = number of observations in the second set of data.

If we have to combine three or more than three sets of data, then the same formula can be generalised as:

$$\bar{X}_{123\ldots} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2 + N_3\bar{X}_3 + \ldots\ldots}{N_1 + N_2 + N_3 + \ldots\ldots}$$

The arithmetic mean has the great advantages of being easily computed and readily understood. It is due to the fact that it possesses almost all the properties of a good measure of central tendency. No other measure of central tendency possesses so many properties. However, the arithmetic mean has some disadvantages. The major disadvantage is that its value may be distorted by the presence of extreme values in a given set of data. A minor disadvantage is. When it is used for open-end distribution since it is difficult to assign a midpoint value to the open-end class.

Activity A

The following data relate to the monthly earnings of 428 skilled employees in a big organisation.

Monthly Earnings (Rs.)	No. of employees	Monthly Earnings (Rs.)	No. of employees
1840-1900	1	2080-2140	126
1900-1960	3	2140-2200	90
1960-2020	46	220O-2260	50
2020-2080	98	2260-2320	6
		2320-2380	8

Compute the arithmetic mean and interpret this value.

7.6 WEIGHTED ARITHMETIC MEAN

The arithmetic mean, as discussed earlier, gives equal importance (or weight) to each observation. In some cases, all observations do not have the same importance. When this is so, we compute weighted arithmetic mean. The weighted arithmetic mean can be defined as

$$\overline{X}_w = \frac{\sum WX}{\sum W}$$

Where $\overline{X}_w$ represents the weighted arithmetic mean,

W are the weights assigned to the variable X.

You are familiar with the use of weighted averages to combine several grades that are not equally important. For example, assume that the grades consist of one final examination and two mid term assignments. If each of the three grades are given a different weight, then the procedure is to multiply each grade (X) by its appropriate weight (W). If the final examination is 50 per cent of the grade and each mid term assignment is 25 per cent, then the weighted arithmetic mean is given as follows:

$$\overline{X}_w = \frac{\sum WX}{\sum W} = \frac{W_1X_1 + W_2X_2 + W_3X_3}{W_1 + W_2 + W_3}$$

$$= \frac{50\ X_1 + 25X_2 + 25X_3}{50+25+25}$$

Suppose you got 80 in the final examination, 95 in the first mid term assignment, as 85 in the second mid term assignment then

$$\overline{X}_w = \frac{50\ (80) + 25\ (95) + 25\ (85)}{100}$$

$$= \frac{4000+2375+2125}{100} = \frac{8500}{100} = 85$$

The following table shows this computation in a tabular form which is easy to employ for calculation of weighted arithmetic mean.

	Grade X	**Weight W**	**WX**
Final Examination	80	50	4000
First assignment	95	25	2375
Second assignment	85	25	2125
		$\sum W = 100$	$\sum WX = 8500$

$$\overline{X}_w = \frac{\sum WX}{\sum W} = \frac{8500}{100} = 85$$

The concept of weighted arithmetic mean is important because the computation is the same as used for averaging ratios and determining the mean of grouped data. Weighted mean is specially useful in problems relating to the construction of index numbers.

Activity B

A contractor employs three types of workers: male, female and children. He pays Rs. 40, Rs. 30, and Rs. 25 per day to a male, female and child worker respectively. Suppose he employs 20 males, 15 females, and 10 children. What is the average wage per day paid by the contractor? Would it make any difference in the answer if the number of males, females, and children employed are equal? Illustrate.

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

7.7 MEDIAN

A second measure of central tendency is the median. Median is that value which divides the distribution into two equal parts. Fifty per cent of the observations in the distribution are above the value of median and other fifty per cent of the observations are below this value of median. The median is the value of the middle observation when the series is arranged in order of size or magnitude. If the number of observations is odd, then the median is equal to one of the original observations. If the number of observations is even, then the median is the arithmetic mean of the two middle observations. For example, if the income of seven persons in rupees is 1100, 1200, 1350, 1500, 1550, 1600, 1800, then the median income would be Rs. 1500. Suppose one more person joins and his income is Rs. 1850, then the median income of eight persons would be $\frac{1500+1550}{2} = 1525$ (since the number of observations is even, the median is the arithmetic mean of the 4^{th} person).

For grouped data, the following formula may be used to locate the value of median.

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

where L is the lower limit of the median class, pcf is the preceding cumulative frequency to the median class, f is the frequency of the median class and i is the size of the median class.

As an illustration, consider the following data which relate to the age distribution of 1000 workers in an industrial establishment.

Age (Years)	No. of workers	Age (Years)	No. of Workers
Below 25	120	40-45	150
25-30	125	45-50	140
30-35	180	50-55	100
35-40	160	55 and above	25

Determine the median age.

The location of median value is facilitated by the use of a cumulative frequency distribution as shown below in the table.

Age (Years)	No. of workers f	Cumulative frequency c.f
Below 25	120	120
25-30	125	245
30-35	180	425
35-40	160	585
40-45	150	735
45-50	140	875
50-55	100	975
55 and Above	25	1000

N = 1000

Median = size of $\frac{N}{2}$ th observation = $\frac{1000}{2}$ = 500th observation which lies in the class 35 - 40.

$$\text{Median} = L + \frac{N/2 - pcf}{f} \times i = 35 + \frac{500 - 425}{160} \times 5$$

$$= 35 + \frac{375}{160} = 35 + 2.34 = 37.34 \text{ years.}$$

Hence the median age is approximately 37 years. This value of median suggests that half of the workers are below the age of 37 years and other half of the workers are above the age of 37 years.

7.8 MATHEMATICAL PROPERTY OF MEDIAN

The important mathematical property of the median is that the sum of the absolute deviations about the median is a minimum. In symbols $\sum|X\text{-Med.}|$ = a minimum.

Although the median is not as popular as the arithmetic mean, it does have the advantage of being both easy to determine and easy to explain.

As illustrated earlier, the median is affected by the number of observations rather than the values of the observations; hence it will be less distorted as a representative value than the arithmetic mean.

An additional advantage of the median is that it may be computed for an open-end distribution.

The major disadvantage of median is that it is a less familiar measure than the arithmetic mean. However, since median is a positional average, its value is not determined by each and every observation. Also median is not capable of algebraic treatment.

Activity C

For the following data, compute the median and interpret this value.

Monthly Rent (Rs.)	No. of Persons paying the rent	Montly Rent (Rs.)	No. of Persons paying the rent
Below 1000	6	1800-2000	15
1000-1200	9	2000-2200	10
1200-1400	11	2200-2400	8
1400-1600	14	2400 and above	7
1600-1800	20		

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

7.9 QUANTILES

Quantiles are the related positional measures of central tendency. These are useful and frequently employed measures of non-central location. The most familiar quantiles are the **quartiles, deciles,** and **percentiles.**

Quartiles: Quartiles are those values which divide the total data into four equal parts. Since three points divide the distribution into four equal parts, we shall have three quartiles. Let us call them Q_1, Q_2, and Q_3. The first quartile, Q_1, is the value such that 25% of the observations are smaller and 75% of the observations are larger. The second quartile, Q_2, is the median, i.e., 50% of the observations are smaller and 50% are larger. The third quartile, Q_3, is the value such that 75% of the observations are smaller and 25% of the observations are larger.

For grouped data, the following formulas are used for quartiles.

$$Q_j = L + \frac{jN/4 - pcf}{f} \times i \qquad \text{for } j = 1,2,3$$

where L is lower limit of the quartile class, pcf is the preceding cumulative frequency to the quartile class, f is the frequency of the quartile class, and i is the size of the quartile class.

Deciles: Deciles are those values which divide the total data into ten equal parts. Since nine points divide the distribution into ten equal parts, we shall have nine deciles denoted by D_1, D2,...................................... , D_9,

For grouped data, the following formulas are used for deciles:

$$D_k = L + \frac{KN/10 - pcf}{f} \times i \qquad \text{for } k = 1, 2, ..., 9$$

where the symbols have usual meaning and interpretation.

Percentiles: Percentiles are those values which divide the total data into hundred equal parts. Since ninety nine points divide the distribution into hundred equal parts, we shall have ninety nine percentiles denoted by

$$P_1, P_2, P_3, ..., P_{99}.$$

For grouped data, the following formulas are used for percentiles.

$$P_l = L + \frac{lN/100 - pcf}{f} \times i \qquad \text{for } l = 1, 2, ..., 99$$

To illustrate the computations of quartiles, deciles and percentiles, consider the following grouped data which relate to the profits of 100 companies during the year 1987-88.

Profits (Rs. lakhs)	No. of companies	Profits (Rs. lakhs)	No. of companies
20–30	4	60–70	15
30–40	8	70–80	10
40–50	18	80–90	8
50–60	30	90–100	7

Calculate Q_1, Q_2, (median), D_6, and P_{90}, from the given data and interpret these values.

To compute Q_1, Q_2, D_6, and P_{90}, we need the following table:

Profits (Rs. lakhs)	No. of companies f	c.f
20–30	4	4
30–40	8	12
40–50	18	30
50–60	30	60
60–70	15	75
70–80	10	85
80–90	8	93
90–100	7	100

Q_1 = Size of N/4th observation $= \frac{100}{4} = 25$th observation, which lies in the class 40 – 50

$$Q_1 = L + \frac{N/4 - pcf}{f} \times i = 40 + \frac{25-12}{18} \times 10 = 40 + 7.22 = 47.22$$

This value of Q_1 suggests that 25% of the companies earn an annual profit of Rs. 47.22 lakh or less.

Median or Q_2 = Size of $\frac{N}{2}$th observation $= \frac{100}{2} = 50$th observation

which lies in the class 50 – 60.

$$Q_2 = L + \frac{N/2 - pcf}{f} \times i = 50 + \frac{50-30}{30} \times 10 = 50 + 6.67 = 56.67$$

This value of Q_2, (or median) suggests that 50% of the companies earn an annual profit of Rs. 56.67 lakh or less and the remaining 50% of the companies earn an annual profits of Rs. 56.67 lakh or more.

D_6 = Size of $\frac{6N}{10}$th observation $= \frac{6x100}{10} = 60$th observation, which lies in the class 50 – 60.

$$D_6 = L + \frac{6N/10 - pcf}{f} \times i = 50 + \frac{60-30}{30} \times 10 = 50 + 10 = 60$$

Thus 60% of the companies earn an annual profit of Rs. 60 lakh or less and 40% of the companies earn Rs. 60 lakh or more.

P_{90} = size of $\frac{90N}{100}$th observation $= \frac{90x100}{100} = 90$th observation, which lies in the class 80 – 90.

$$P_{90} = L + \frac{90N/100 - pcf}{f} \times i = 80 + \frac{90-85}{10} \times 10 = 80 + 5 = 85$$

This value of 90th percentile suggests that 90% of the companies earn an annual profit of Rs. 85 lakh or less and 20% of the companies earn more than Rs. 85 lakh or more.

7.10 LOCATING THE QUANTILES GRAPHICALLY

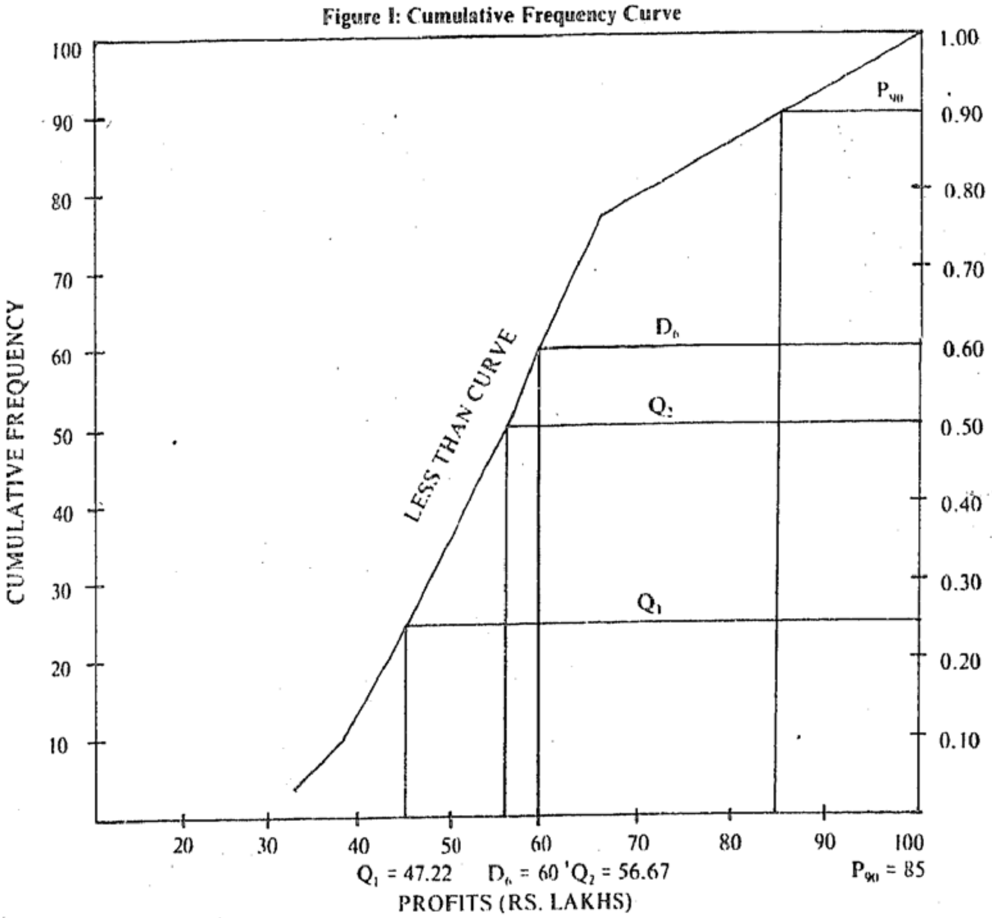
To locate the median graphically, draw less than cumulative frequency curve (less than ogive). Take the variable on the X-axis and frequency on the Y-axis. Determine the median value by locating N/2th observation on the Y-axis. Draw a horizontal line from this on the cumulative frequency curve and from where it meets the curve, draw a perpendicular on the X-axis. The point where it meets the X-axis is the value of median.

Similarly we can locate graphically the other quantiles such as quartiles, deciles and percentiles.

For the data of previous illustration, locate graphically the values of Q_1, Q_2, $D_{60,}$ and Q_{90}.

The first step is to make a less than cumulative frequency curve as shown in figure I.

Figure I: Cumulative Frequency Curve



To determine different quantiles graphically, horizontal lines are drawn from the cumulative relative frequency values. For example if we want to determine the value of median (or Q_2), a horizontal line can be drawn from the cumulative frequency value of 0.50 to the less than curve and then extending the vertical line to the horizontal axis. Ina similar way, other values can be determined as shown in the graph. From the graph, we observe

$Q_1 = 47.22$, $Q_2 = 57.67$, $D_2 = 60.0$, $P_{90} = 85$

It may be noted that these graphical values of quantiles are the same as obtained by the formulas.

Activity D

Given below is the wage distribution of 100 workers in a factory:

Wages (Rs.)	No. of workers	Wages (Rs.)	No. of workers
Below 1000	3	1800-2000	10
1000-1200	5	2000-2200	8
1200-1400	12	2200-2400	5
1400-1600	23	2400 and above	3
1600-1800	31		

Draw a less than cumulative frequency curve (ogive) and use it to determine graphically the values of Q_2, Q_3, D_{60}, and P_{80}. Also verify your result by the corresponding mathematical formula.

…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………

7.11 MODE

The mode is the typical or commonly observed value in a set of data. It is defined as the value which occurs most often or with the greatest frequency. The dictionary meaning of the term mode is most usual'. For example, in the series of numbers 3, 4, 5, 5, 6, 7, 8, 8, 8, 9, the mode is 8 because it occurs the maximum number of times.

The calculations are different for the grouped data, where the modal class is defined as the class with the maximum frequency. The following formula is used for calculating the mode.

$$\text{Mode} = L + \frac{d_1}{d_1 + d_2} \times i$$

where L is lower limit of the modal class, d_1 is the difference between the frequency of the modal class and the frequency of the preceding class, d_2 is the difference between the frequency of the modal class and the frequency of the succeeding class, i is the size of the modal class. To illustrate the computation of mode, let us consider the following data.

Daily Sales (Rs. thousand)	No. of firms	Daily Sales (Rs. thousand)	No. of firms
20-30	15	60-70	35
30-40	23	70-80	25
40-50	27	80-90	5
50-60	20		

Since the maximum frequency 35 is in the class 60-70, therefore 60-70 is the modal class. Applying the formula, we get

$$\text{Mode} = L + \frac{d_1}{d_1 + d_2} \times i = 60 + \frac{35 - 20}{(35 - 20) + (35 - 25)} \times 10$$

$$= 60 + \frac{150}{25}$$

$$= 60 + 6 = \text{Rs.}66.$$

Hence modal daily sales are Rs. 66.

7.12 LOCATING THE MODE GRAPHICALLY

In a grouped data, the value of mode can also be determined graphically. In graphical method, the first step is to construct histogram for the given data. The next step is to draw two straight lines diagonally on the inside of the modal class bars, starting from each upper corner of the bar to the upper corner of the adjacent bar. The last step is to draw a perpendicular line from the intersection of the two diagonal lines to the X-axis which gives us the modal value.

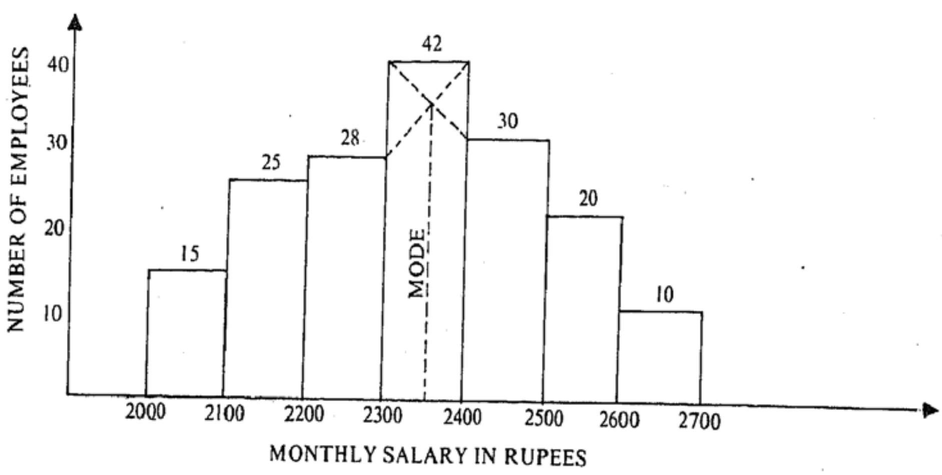
Consider the following data to locate the value of mode graphically.

Monthly salary (Rs.)	No. of employees	Monthly salary (Rs.)	No. of employees
2000-2100	15	2400-2500	30
2100-2200	25	2500-2600	20
2200-2300	28	2600-2700	10
2300-2400	42		

First draw the histogram as shown below in figure II.

Figure II: Histogram of Monthly Salaries

Figure II: Histogram of Monthly Salaries



The two straight lines are drawn diagonally in the inside of the modal class bars and then finally a vertical line from the intersection of the two diagonal lines is drawn on the X-axis. Thus the modal value is approximately Rs. 2353. It may be noted that the value of mode would be approximately the same if we use the algebric method.

The chief advantage of the mode is that it is, by definition, the most representative value of the distribution. For example, when we talk of modal size of shoe or garment, we have this average in mind. Like median, the value of mode is not affected by extreme values and its value can be determined in open-end distributions.

The main disadvantage of the mode is its indeterminate value, i.e., we cannot calculate its value precisely in a grouped data, but merely estimate it. When a given set of data have two or more than two values as maximum frequency, it is a case of bimodal or multimodal distribution and the value of mode cannot be determined. The mode has no useful mathematical properties. Hence, in actual practice the mode is more important as a conceptual idea than as a working average.

Activity E

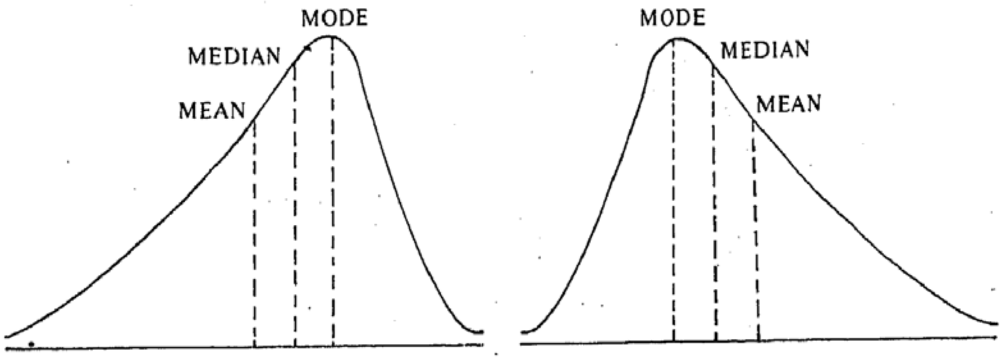
Compute the value of mode from the grouped data given below. Also check this value of mode graphically.

Monthly stipend (Rs.)	No. of management trainees	Monthly stipend (Rs.)	No. of trainees
2500-2700	25	3300-3500	20
2700-2900	35	3500-3700	15
2900-3100	60	3700-3900	5
3100-3300	40		

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

7.13 RELATIONSHIP AMONG MEAN, MEDIAN AND MODE

A distribution in which mean, median and mode coincide is known as a **symmetrical** (bell shaped) distribution. If a distribution is skewed (that is, not symmetrical) then mean, median, and mode are not equal. In a moderately skewed distribution, a very interesting relationship exists among mean, median and mode. In such type of distributions, it can be proved that the distance between mean and median is approximately one third of the distance between the mean and mode. This is shown below for two types of such distributions.



This relationship can be expressed as follows:

Mean - Median = 1/3 (Mean - Mode)

or Mode = 3 Median - 2 Mean

Similarly, we can express the approximate relationship for median in terms of mean and mode. Also this can be expressed for mean in terms of median and mode. Thus, if we know any of the two values of the averages, the third value of the average can be determined from this approximate relationship.

For example, consider a moderately skewed distribution in which mean and median is 35.4 and 34.3 respectively. Calculate the value of mode.

To compute the value of mode, we use the approximate relationship

Mode □ 3 Median - 2 Mean

= 3 (34.3) - 2 (35.4)

= 102.9-70.8 = 32.1

Therefore the value of mode is 32.1.

7.14 GEOMETRIC MEAN

The geometric mean like the arithmetic mean, is a calculated average. The geometric mean, GM, of a series of numbers, X_1 X_2, ... X_n, is defined as

$$GM = \sqrt[N]{X_1.X_2.X_3..........X_N}$$

or the N^{th} root of the product of N observations.

When the number of observations is three or more, the task of computation becomes quite tedious. Therefore a transformation-into logarithms is useful to simplify calculations. If we take logarithms of both sides, then the formula for GM becomes

$$\text{Log GM} = \frac{1}{N}\ (\text{loog } X_1 + \log X_2 + ...+ \log X_N)$$

$$GM = \text{Antilog}\left(\frac{\sum \log X}{N}\right)$$

$$\text{and therefore, } GM = \text{Antilog}\left(\frac{\sum \log X}{N}\right)$$

For the grouped data, the geometric mean is calculated with the following formula

$$GM = \text{Antilog}\left(\frac{\sum f \log X}{N}\right)$$

Where the notation has the usual meaning.

Geometric mean is specially useful in the construction of index numbers. It is an average most suitable when large weights have to be given to small values of observations and small weights to do large values of observations. This average is also useful in measuring the growth of population.

The following data illustrates the use and the computations involved in geometric mean.

A machine was purchased for Rs. 50,000 in 1984. Depreciation on the diminishing balance was charged @ 40% in the first year, 25% in the second year and 15% per annum during the next three years. What is the average depreciation charged during the whole period?

Since we are interested in finding the average rate of depreciation, geometric mean will be the most appropriate average.

Year	Diminishing value (for a value of Rs. 100) X	Log X
1984	100 – 40 = 60	1.77815
1985	100 – 25 = 75	1.87506
1986	100 – 15 = 85	1.92941
1987	100 - 15 = 85	1.92941
1988	100 – 15 = 85	1.92941
		$\sum \log X = 9.44144$

$$GM = \text{Antilog}\left(\frac{\sum \log X}{N}\right)$$

$$= \text{Antilog}\left(\frac{9.44144}{5}\right) = \text{Antilog } 1.8883 = 77.32$$

The diminishing value being Rs. 77.32, the depreciation will be 100-77.32 = 22.68%.

The geometric mean is very useful in averaging ratios and percentages. It also helps in determining the rates of increase and decrease. It is also capable of further algebraic treatment, so that a combined geometric mean can easily be computed.

However, compared to arithmetic mean, the geometric mean is more difficult to compute and interpret. Further, geometric mean cannot be computed if any observation has either a value zero or negative:

Activity F

Find the geometric mean for the following data:

Class interval	Frequency	Class interval	Frequency
4.5 – 5.5	8	8.5 – 9.5	25
5.5 – 6.5	10	9.5 – 10.5	18
6.5 – 7.5	12	10.5 – 11.5	7
7.5 – 8.5	15	11.5 – 12.5	5

…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………

7.15 HARMONIC MEAN

The harmonic mean is a measure of central tendency for data expressed as rates such as kilometers per hour, tonnes per day, kilometers per litre etc. The harmonic mean is defined as the **reciprocal of the arithmetic mean of the reciprocal** of **the individual observations.** If $X_1, X_2, \cdots \cdots\cdots \ldots X_N$ are N observations, then harmonic mean can be represented by the following formula.

$$HM = \frac{N}{\frac{1}{X_1}+\frac{1}{X_2}+\ldots\ldots+\frac{1}{X_N}} = \frac{N}{\sum\left(\frac{1}{X}\right)}$$

For example, the harmonic mean of 2, 3, 4 is

$$HM = \frac{3}{\frac{1}{2}+\frac{1}{3}+\frac{1}{4}} = \frac{3}{13/12} = \frac{36}{13} = 2.77$$

For grouped data, the formula becomes

$$HM = \frac{N}{\sum\left(\frac{f}{X}\right)}$$

The harmonic mean is useful for computing the average rate of increase of profits, or average speed at which a journey has been performed, or the average price at which an article has been sold. Otherwise its field of application is really restricted.

To explain the computational procedure, let us consider the following example.

In a factory, a unit of work is completed by A in 4 minutes, by B in 5 minutes, by C in 6 minutes, by D in 10 minutes, and by E in 12 minutes. Find the average number of units of work completed per minute.

The calculations for computing harmonic mean are given below:

X	1/X
4	0.250
5	0.200
6	0.167
10	0.100
12	0.083
	$\sum 1/X =$ 0.8

Hence the average number of units computed per minute is 6.25.

The harmonic mean like arithmetic mean and geometric mean is computed from each and every observation. It is specially useful for averaging rates.

However, harmonic mean cannot be computed when one or more observations have zero value or when there are both positive or negative observations. In dealing with business problems, harmonic mean is rarely used.

Activity G

In a factory, four workers are assigned to complete an order received for dispatching 1400 boxes of a particular commodity. Worker-A takes 4 minutes per box, B takes 6 minutes per box, C takes 10 minutes per box, D takes 15 minutes per box. Find the average minutes taken per box by the group of workers.

………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………

7.16 SUMMARY

Measures of central tendency give one of the very important characteristics of data. Any one of the various measures of central tendency may be chosen as the most representative or typical measure. The arithmetic mean is widely used and understood as a measure of central tendency. The concepts of weighted arithmetic mean, geometric mean, and harmonic mean are useful for specified type of applications. The median is generally a more representative measure for open-end distribution and highly skewed distribution. The mode should be used when the most demanded or customary value is needed.

7.17 KEY WORDS

Arithmetic Mean is equal to the sum of the values divided by the number of values.

Geometric Mean of N observations is the Nth root of the product of the given value observations.

Harmonic Mean of N observations is the reciprocal of the arithmetic mean of the reciprocals of the given values of N observations.

Median is that value of the variable which divides the distribution into two equal parts.

Mode is that value of the variable which occurs the maximum number of times.

Quantiles are those values which divide the distribution into a fixed number of equal parts, eg., quartiles divide distribution into four equal parts.

7.18 SELF-ASSESSMENT EXERCISES

1 List the various measures of central tendency studied in this unit and explain the difference between them.

2 Discuss the mathematical properties of arithmetic mean and median.

3 Review for each of the measure of central tendency, their advantages and disadvantages.

4 Explain how you will decide which average to use in a particular problem.

5 What are quantiles? Explain and illustrate the concepts of quartiles, deciles and percentiles.

6 Following is the cumulative frequency distribution of preferred length of study-table obtained from the preferency study of 50 students.

Length	No. of students	Length	No. of students
more than 50 cms	50	more than 90 cms	25
more than 60 cms	46	more than 100 cms	18
more than 70 cms	40	more than 110 cms	7
more than 80 cms	32		

A manufacturer has to take decision on the length of study-table to manufacture. What length would you recommend and why?

7 A three month study of the phone calls received by Small Company yielded the following information.

Number of calls per day	No. of days	Number of calls per day	No. days
100 - 200	3	600 - 700	10
200 - 300	7	700 - 800	9
300 - 400	11	800 - 900	8
400 - 500	13	900 - 1000	4
500 - 600	27		

Compute the arithmetic mean, median and mode.

From the following distribution of travel time of 213 days to work of a firm's find the modal travel time.

Travel time (in minutes)	No. of days	Travel time (in minutes)	No. of days
Less than 80	213	Less than 40	85
Less than 70	210	Less than 30	50
Less than 60	195	Less than 20	13
Less than 50	156	Less than 10	2

9 The mean monthly salary paid to all employees in a company is Rs. 1600. The mean monthly salaries paid to technical employees are Rs. 1800 and Rs. 1200 respectively. Determine the percentage of technical and non-technical employees of the company.

10 The following distribution is with regard to weight (in grams) of apples of a given variety. If an apple of less than 122 grams is to be considered unsuitable for export, what is the percentage of total apples suitable for the export?

Weight (in grams)	No. of apples	Weight (in grams)	No. of apples
100-110	10	140-150	35
110-120	20	150-160	15
120-130	40	160-170	5
130-140			

Draw an ogive of more than one type and deduce how many apples will be more than 122 grams.

11 The geometric mean of 10 observations on a certain variable was calculated to be 16.2. It was later discovered that one of the observations was wrongly recorded as 10.9 when in fact it was 21.9. Apply appropriate correction and calculate the correct geometric mean

12 An incomplete distribution of daily sales (Rs. thousand) is given below. The data relate to 229 days.

Daily sales (Rs. thousand)	No. of days	Daily sales (Rs. thousand)	No. of days
10-20	12	50-60	?
20-30	30	60-70	25
30-40	?	70-80	18
40 -50			

You are told that the median value is 46. Using the median formula, fill up the missing frequencies and calculate the arithmetic mean of the completed data.

13 The following table shows the income distribution of a company.

Income (Rs.)	No. of employees	Income (Rs.)	No. of employees
1200-1400	8	2200-2400	35
1400-1600	12	2400-2600	18
1600-1800	20	2600-2800	7
1800-2000	30	2800-3000	6
2000-2200	40	3000-3200	4

Determine (i) the mean income (ii) the median income (iii) the mean (iv) the income limits for the middle 50% of the employees (v) D_7, the seventh docile, and (vi) P_{80}, the eightieth percentile.

7.19 FURTHER READINGS

Clark, T.C. and E. W. Jordan, 1985. *Introduction to Business and Economic Statistics,* South-Western Publishing Co.

Enns, P.G., 1985. *Business Statistics.* Richard D. Irwin: Homewood.

Gupta, S.P. and M.P. Gupta, 1988. *Business Statistics,* Sultan Chand & Sons: New Delhi.

Moskowitz, H. and G.P. Wright, 1985. *Statistics for Management and Economics,* Charles E. Merin Publishing Company: