PUNJABI UNIVERSITY PATIALA

B.COM.–II (SEMESTER-III)

B.C. - 306

BUSINESS STATISTICS

UNIT NO. : I

BOTH MEDIUM

Lesson No.:

UPDATED ON: JULY 05, 2023

(2021-22, 2022-23, 2023-24)

BC 306: BUSINESS STATISTICS

Time allowed : 3 hours Max Marks: 100

Pass Marks : 35% Internal Assessment: 30

Periods per week : 6 External Assessment: 70

Instructions for Paper-Setters/Examiners

The question paper covering the entire course shall be divided into three sections as follows:

SECTION-A

It will consist of essay type questions. Four questions (two theory and two numericals) shall be set by the examiner from Unit-I of the syllabus and the candidate shall be required to attempt two. Each question shall carry 10 marks; total weight of the section shall be 20 marks.

.

SECTION-B

It will consist of essay type questions. Four questions shall (two theory and two numericals) be set by the examiner from Unit-II of the syllabus and the candidate shall be required to attempt two. Each question shall carry 10 marks; total weight of the section shall be 20 marks.

.

SECTION-C

It will consist of 12 very short answer questions (six theory and six numericals) from entire syllabus. Students are required to attempt 10 questions up to five lines in length. Each question shall carry 3 marks; total weight of the section shall be 30 marks

UNIT-I

Introduction to Statistics-Definition, Importance and Limitations, Functions and scope Measures of Central Tendency: Mean, Median, Mode. Measures of dispersion: Range, Quartile deviation, Mean deviation and Standard deviation.

Analysis of Time Series: Causes of variations in time series multiplicative models; Determination of trends, Moving averages method and method of least squares (including linear, second degree, parabolic and exponential trends); Computation of seasonal-indices by simple averages, ratio-trend, ratio-to-moving average, and link relative methods.

UNIT-II

Index numbers: Need, definition and limitations of index numbers-simple and weighted index numbers- Laspyer's, Paasche's and Fisher Index numbers, Criterion of ideal index numbers, problems involved in the construction of index numbers.

Correlation: Meaning, types and measurement of correlation (Karl Pearson's methods and Spearman's rank correlation).

Regression: Meaning, Regression Equation of X on Y and Y on X.

Forecasting Methods: Forecasting Concept, types and importance; General approach to forecasting; Methods of forecasting; Forecasting demand; Industry Vs. Company sales forecasts; Factors affecting company sales.

Course Outcome: After studying this course, students will acquire knowledge on the application of statistical techniques for data analysis. This knowledge can be used by them in their research projects.

Books Recommended

1. R.P. Hooda : Statistics for Business and Economics
2. S.P. Gupta : Statistics Methods
3. S.C. Gupta and V.K. Kapoor : Fundamentals of Applied Statistics.

B.COM.–II (SEMESTER-III) B.C. 306

BUSINESS STATISTICS

LESSON NO. 1.1 AUTHHOR: DR. AMANPREET KAUR

INTRODUCTION TO STATISTICS

1.1.1 Objective

1. The main objective of this lession to understand the meaning and importans of statistics.

2. To understand the function of statistics.

1.1.2 Introduction

Growing complexities of human natural/Social Phenomenon has left no alternative to the world decision makers but to depend upon the cordinal as well as ordinal methods of measurement and interpretation of the varied problems. Nearly, every activity-mental or physical or natural, is measured and interpreted quantitatively. Today whole world lives in the world of numbers/Statistic and reaches to conclusive decisions on the basis of statistical knowledge. Perhaps this may be the reason that Solomon Fabricant visualised the prospects of development of statistical use in these words. "Increasing public interest in and demand for social statistics rests on the basic premise that th problem of the Society, as well as natural science and Technology, can be solved by the increase and diffusion of this especially matter-of-fact type of matter-of-fact knowledge. The whole world now seems to hold that statistics can be useful in understanding, assessing and controlling the operations of Society." Also, the forecast H.G. Wells that "Statistical thinking wil one day be as the ability, to read and write", seems to prove true. Today the authencity of your statements is suspected if you have not expressed it numerically. "When you can measure what you are speaking about and express it in numbers, your knowledge is of a meagre and unsatisfactory kind". In short, statistical way of thought has entered the whole arena of human thinking in such a way that one becomes curious to know about its origin and growth.

1.1.3 Meaning and Definition of Statistics

(a) *The Meaning:* Statistics, like many words have different meanings in different context. Some people regard statistics as data, facts or measurement while others believe it to be the study of figures. There are another group of people who consider it as analysis of figures for forecasting or drawing inferences. besides this, the representation of facts in the form of diagrams a graph or maps is also supposed to be statistics. Processing , analysis and application of quantitative facts is regarded as statistics. In this way statistics imbibes three forms.

(i) *As Numerical Data:* In the product form, it represents the numerical data, such as statistics of national income, unemployment, imports and exports, etc. It is used here in plural form.

(ii) *As a Subject:* In the process form, it is used as subject and is in Singular form, like Economics, Physics etc. In this sense, the term 'Statistics' refers the whole field of study of which 'Statistics' in the plural sense are the subject-matter. In other words if refers to the statistical principles and methods used in collection, analysis and interpretation of data. These methods finally help in taking decisions and testing the hypothesis.

(iii) *In its modern connotation:* it may also refer to the study of and research into the theory and principle underlying statistical methods. It is the field of study that pioneers in, expands the frontiers of statistical methodology and uses.

(iv) "Statistics" is also used by the experts in the field, for the terms like mean, median, mode, standard deviations, etc. calculated from sample.

In brief statistics is used both in singular and plural form. It is also used to represent (i) Numerical data, (ii) Statistical science Subject, and (iii) statistical measures, which is clear from Tate's this statement 'You compute statistics, by statistics from statistics."

1.1.4 Nature & Scope of Statistics

The Nature: As we have seen above that some statisticians have put statistics in the category of science, while others believe it to be Arts. There are many members who believe it to the both Art and Science. To decide the nature of statistics it has to be examined in both the categories.

(a) *Statistics as a Science:* "Science is a body of systematized knowledge". In this way, any subject can be put in the category of science if it possess following characteristics:

(i) It is a systematised group of knowledge.
(ii) Its laws and methods must be Universally acceptable.
(iii) It must analyse the cause-effect relationship.
(iv) It must possess the quality of estimation and forecasting.

1.1.5 The Function of Statistics

Statistics performs many functions useful to human beings. Robert W. Buges elaborates the functions of statistics in these words. "The fundamental gospect of statistics is to push back the domain of ignorance, prejudice, rule of thumb, arbitrary and premature decisions, tradition and dogmatism and to increase the domain in which decisions are made and principles are made on the basis of analysed quantitative facts". the evergrowing popularity as a quantitative methods is because of the functions, which statistics performs:

(i) *Statistics provides definiteness to the facts:* Quantitative facts can easily be believed and trusted in comparison to abstract and qualitative facts. Statistics summarises the generalised facts and presents them in a definite form. Various characteristics pertaining to some phenomena, become easily understandable, if they are expressed in numbers. For instance, the index Number technique expresses the complex variables into a form, which can easily be understood, It is easy to understand that prices index of consumer items has gone up by 10 per cent, instead of saying that prices are increasing leaps and bounds.

1.1.6 Simplifies the complexities

It is very difficult for an individual to understand and conclude from huge numerical data. Statistical methods try to present the great mass of complex data into simple and understandable form. For example, statistical technique like mean, median, variation, correlation, graphs and diagram etc. make the complex data intelligible and understandable in short period and in better way. W.L. King defining the function of statistical science rightly wrote, "It is for the purpose of simplifying these unwidely masses of facts that statistical science is useful. It reduces them to numerical totals or average which may be abstractly handled like any other more numbers. It draws pictures and diagram to illustrate general tendencies and thus in many ways adapts these group of ideas to the capacity of our intellects."

1.1.7 The Function of Statisticians

A statistician is a person who collects the data with the help of Statistical techniques for some definite purpose of enquiry, analysis and interprets the facts as they are. He is in other words, practitioner of Art & Science of statistics. Rhode divides the functions of three parts. "In the first place he is concerned with the assembling of statistical data, in the second place with their analysis, and in the third place with the interpretation of the results of such an analysis."

For the sake of convenience, the functions of a statistician can be described in four categories:

(i) *The Observation:* This is the first and more most important function of a statistician. In the beginning, he ponders over the objectives of research and plans the enquiry after deep thinking about the time-schedule, economics situation and available resources, the area and scope of research, time involved, level of precision and accuracy desire, and modallilies of data collection. It is also his duty to specify and decide the manpower required, suited to his needs. All these functions, which seems to be primary and ordinary, requires a thorough skill and expertise based on wide observation experience. This should planned with great care and confidence.

(ii) *The Collection:* The collection of data with pre-determined and pre-planned method is second important function of a statistician. It is he, who decides the method of collection of data, whether it should be through enumeration or through estimation? He collects both-primary and secondary data for the purpose. To test the authenticity and accuracy of collected data, he edits them also and presents them in tabular form for the purpose of analysis and interpretation.

(iii) *The Analysis:* The analysis work of statistician is very wide and cumbersome. He has to perform many works like classifying, serialising and bring. them to a comparable form. After this, he has to compute different statistical parameter like average, standard deviation, skewness, regression coefficient, correctional coefficients, coefficient of associations, etc., for establishing relationship between them.

1.1.8 Importance of Statistics

(1) Scope in Social Sciences :- Statistics help the researcher to use the information in comprehensive manner to describe and predict the trend and behavioural pattern, when both Quantitative and Qualitative information are used in social sciences. Social Sciences studies the characteristics of the population assuming normal distribution and the important decisions regarding variables under study are taken to explain the behaviour of the activities by using either parametric or non parametric statistical tests.

(2) Scope for Education :- Education remained primarily on state agenda particularly in developing and under developed countries. But for proper expansion of education in the better interest of the society, the characteristic of the students, contents of the teacher and infrastructural facilities should be properly defined. Statistics

caters to the need of the society and nation by analysing these characteristics of students and contents of the teacher through describing the parameters of all the components.

(3) Scope for solving the problems :- Statistics makes it possible to find out best possible solution to a problem, when there is useful difference between two or more variables by using statistical tools like dispersion correlation and regression. Statistics helps the individual investigator to analysis all the alternative solutions and to choose best out of them, while minimizing the error factor.

(4) Scope for commerce :- All the social systems evolved in the ancient times have been transformed into different economic systems as a result of involvement of money matters in almost all the spheres. Therefore, it has become important to handle the funds properly mobilising in various sectors of the economy. The study of cost and benefit analysis helps to take vital decisions, regarding the best option of making investment, mobilising financial assets and liquidity of the assets, where and when they yield maximum return.

(5) Scope for formulating the theories :- Theories regarding the various phenomenon's of social, religious and economic life are formulated on the basis of statistical facts collected from the various fields. Investigators use factual data and various statistical tools to check whether a particular theory can be proved and made applicable for maintaining the social and economic laws in the society and in a particular economic system. Statistics helps the investigator to explore the connectivity among various significant facts and factors.

(6) Statistics and Business :- A good businessman makes a good sales prediction, keeping in view all types of seasonal and market variations. The big industrial units test the quality of raw materials they purchase and the finished goods they sell by methods of sampling. The businessmen can also study the additional requirements which propels people to like their products. The businessmen give sale-incentives for their agencies in the form of higher commissions, price-cuts etc. It is possible only if relevant data are available. Thus in modern business, it is due to availability of data, that suitable policies can be followed in case of production, investment, marketing and sale management.

(7) Statistics and Mathematics :- According to F.C. Mills, "Statistics is an offspring of Mathematics". Both are very closely in touch with each other ever since the 17th century. Both have helped jointly in the development of other social and physical sciences. Law of Inertia of Large Numbers' or 'Law of Statistical Regularity' which is the basis of modern theory of statistics are formulated in the Mathematical Theory of Probabilities. Mathematical Averages, co-efficients, graphical representations are used

in containing statistical data. Algebra is very helpful in the field of statistics. Mathematics has helped in the growth and development of various economic theories having links with statistics.

(8) Statistics and Economics :- Prof. Alfred Marshall in the year 1890 observed that "Statistics are the straw out of which I, like every Economist have to make bricks." This statement sums up the importance of statistics in economics. In Economics, all problems are concerned with statistics. Every branch of Economics takes support from statistics in order to prove various Economic theories in it. For instance, if we want to study the expenditure pattern of the people, we can collect the information from the consumers. How much income is spent on different heads of consumption can be known by collecting relevant information.

In the field of production, we can study the effect of certain incentives on production. Again, it is with the help of statistical techniques that we can know the effectiveness of economic policies on production.

In case of trade both domestic and foreign, it is the statistics of prices and costs which play an important role.

The modern states have followed the 'socialistic pattern of society'. It means equal distribution of income and wealth. This is possible if data pertaining to income and wealth is collected. This will help us in assessing whether inequalities are on the increase or decline.

Furthermore, statistics is the most significant tool for the purpose of research in Economics.

(9) Statistics in Insurance :- Statistical information is the backbone of insurance companies. Insurance companies completely depend upon the formation of premium tables on the basis of which they charge their customers for proving risk coverage. The companies take into consideration the age structure tables and then taking life expectancy on the basis of the theory of probability, determine the rates of premium which are generally low in case of younger generation and high in case of upper age bracket. All this movement along the ladder of high or low premium derives its strength from the study of statistics.

(10) Statistics in Planning :- The less-developed or undeveloped countries can grow in short period only if they have a sound planning system. We need planning at all levels. Making good plans does not serve the purpose. Our plans must be fully supported by adequate statistics. Inadequacy of data has proved to be a big problem in planning. So sound planning should be based on sound data. National income is produced, but what is the contribution of Agriculture, Industry, Transport, Trade and

other services, can be known only if we collect data. This will also determine how much finds are to be allocated to different sectors. Not only that, even all predictions are possible only with the help of statistics.

(11) Statistics and Econometrics :- Econometrics is a recent subject. It studies the application of statistical techniques to the economic methods. The study of consumption function, production function, exchange and distribution have been possible with the help of statistical methods. Econometrics, as a subject, is gaining significance. This subject is now called Research Methodology in Economics.

(12) Statistics and Physical and Natural Sciences :- In the physical sciences like Astronomy, Geology and Physics, the statistical methods are used. But in the recent times, the importance of statistics has further increased due to added importance of natural sciences like Medicines, Meteorology, Zoology, Botany etc. Prof Karl Pearson used statistical methods in Biology. He showed that doctrine of evolution and heredity is based on statistical knowledge. The contention that tall fathers have, in general, tall sons can be provided by statistical data. In the end, we can say that it is difficult to find any scientific activity where statistical data and statistical methods are not used.

Thus from the above discussion, it is clear that statistics are important and are applicable anywhere and everywhere. Statistics are used in physical sciences, and social sciences. The significance of statistics has increased from the 'science of the kings' to the science of universal applicability. Statistics is used in all branches of knowledge whether exact or inexact.

1.1.9 Limitations of Statistics

Even though statistics have served the mankinds in many ways and in many front from peace to war and is being utilised by almost every field of knowledge for its advancement and further researches, it is not free from shortcomings which restrict its scope and usefulness. It is always advisable to use it keeping its limitations in mind. Newsholms cautions about its limitations in these words "it must be regarded as an instrument of research of great value, but having several limitations, which are not possible to overcome and as such they need our careful attention". Tippett has also suggested to be careful in the use of stastics in these words, "The application of statistical methods of investigations in the technological and indeed in any other field is based on assumptions, is subjectto limitations and often leads to uncertain results." These limitations are:

(i) *Statistics fails to study qualitative phenomenon:* Science of statistics, as discussed, deals with a set of numerical data, and can be applied to the study of only those phenomena, which can be expressed in numbers like qualitatively or in num- bers. But besides this, those facts which are not measurable/expressed in numbers like beauty, honesty, appreciation, intelligence, health, eagerness, etc. can not be studied unless these qualities/virtues/attributes are reduced into precise quantitative terms. Prof. Horace Secrist wrote this in these words. "Some phenomena can not be quantitatively measured, honesty of resourcefulness, integrity, goodwill, all important in industry, as well as in life, generally, are not susceptible of direct statistical measurement". However, some statistical techniques like Analysis of At- tributes, scaling techniques, weightage technique etc. can be used to qualitative phe- nomena indirectly. If they are assigned certain numerical scores/weightage. For ex- ample, efficiency and efficacy of Malaria Eradication Programme can be studied by teh number of persons saved from death suffering from malaria. Intelligence of a person can be studied by the marks obtained in a particular examination. Liking and disliking can be studied with the help of Projective Techniques. Impact of he- redity can be studied by coefficient of association or coefficient of collignation. All these measurement techniques will be indirect aspect of the phenomena and they have to be expressed quantitatively. In this way it limits the scope of applicability of statistical techniques to the study of quantifiable variables only.

1.1.8 Distrust of Statistics

There can not be two opinions about the utility and applicability of statistics for human welfare. This would enhance in future provided it gives trustworthy, ac- curate and valid inferences. The credibility of statistics is being questioned day-by- day. Public distrust is increasing. In other words, public loses its belief, faith and confidence in the science of statistics and starts condemning it. This distrust has cropped up owing to improper use of statistical tools by unscrupulous, irresponsible, in experienced and dishonest persons having no expertise in the field. This has been expressed by many experts in the field in different ways. Some people call it, "as a tissue of falsehood", Mark Twain once remarked that "The there are three de- grees of comparison in lying, lies, clammed lies, and statistics", and treats statistics as the superlative degree of lying. According La Gordia, wrapped statistics is better than Hitler's Big Lie, it misleads, yet it can not by pinned on you." People say than an ounce of truth will produce tons of statistics, or statistics are lies of the first order. It has ben remarked that, there are black lies, white lies, multichromatic lies', statis- tics is a rainbow of lies.

However, the general conviction that "Statistics can prove any thing, or what statistics reveal in ordinary, but what they write is vital; etc., compels an ordinary person to regard a statistician as naive, incautious and something of pseudo-magician. These statements clearly indicate the extent to which the science of statistics had come to disrepute.

However, there are people who have contrary opinions. The power of science of statistics is great. It can prove anything. There are always tow aspects of an event. It depends upon you how you look. But one thing is clearly true fault does not lie in data but in the technique how it is used

.

1.1.9 Summary

It appears that the adage "Statistical thinking will one day be as the ability, to read and write," is accurate. If you don't articulate your statements mathematically today, they are questioned as to their veracity. In a nutshell, the word "statistics" can be used singly or plurally. Tate's comment that "You compute statistics, by statistics from statistics" makes it obvious that it is also used to represent numerical data, statistical science subjects, and statistical measures.

1.1.10 Glossary

Statistics: The science of collecting, organizing, analyzing, interpreting, and presenting data.

Data: Facts, numbers, or information collected for analysis. It can be qualitative (categorical) or quantitative (numerical).

1.1.11 Self check exercise

Explain population?

Explain parameter?

1.1.11 Short question

descriptive Statistics

Quantitative Variable

1.1.12 Exercise Set

1. 'Statistical methods are most dangerous tools in the hands of the inexperts'. Explain fully the significance of the above statement.
   (B.Com Merrut, 1972, M.Com Agra 1964, M.A. Agra 1973, M.A. Sagar 1987)
2. "Figures do not lie." "Statistics can prove anything". Explain and Reconcile the two statements.
   (B.Com Vikram, 1969, Gorakh, 1969)
3. (a) What is statistics? Discuss its scope and limitations. (b) Write an essay on:

"Statistics in the service of Trade and Commerce."

(B.Com. Punjab, 1973)

4. What are the shortcoming of statistics? Can these shortcomings be overcome? (B.Com. Raj., 1973)

5. Describe with the help of suitable illustrations the functions of statistics. (B.Com. Meerut, 1970; Jiwajee, 1989)

6. "The proper function of statistics, indeed, is to enlarge, individual experi- ence". Comment on the above statement and also explain the functions of a statistician. (B.Com. Agra, 1964)

7. Explain how in modern age statistics can be treated as the science of human welfare. (B.Com. Vik. 1971, 1989)

8. Write an essay on "The role of the Statistician in contemporary society". (M.A. Punjab, 1969)

9. Write an essay on "Statistics in the service of State". (B.Com., 89)

10. 'Statistics are the strawout of which like every other economist, have to make bricks'. (Marshall). Elucidate this statement and indicate the utility of Statis- tics in Economic Planning in India. (B.Com. Kanpur)

11. "Statistics arose from practical requirements of problems in various shapers and its importance is due to its uses in treating such problems". Discuss giving suitable example. (M.A. Meerut, 71)

12. Planning on the basis of inadequate and inaccurate statistics is worse than no planning at all." (Third Five Year Plan). Explain this statement and discuss the importance of statistics in the planned economic development of India.

(B.Com. Vikram, 1968-89)

13. "Planning without statistics is a ship without rudder and compass". In the light of this statement, explain the importance of statistics as an effective aid to national planning in India.

(B.Com. Vikram, 1990., Agra M.Com. 1979)

14. "Statistics plays an important part not only in the study of Economics and Commerce, but also in actual business". Explain fully.

(Raj. B.Com. (T.D.C., 1971)

15. Waht is statistics? Explain the importance of statistical.

16. "A knowledge of statistics is like the knowledge of foreign language or of algebra. It may prove of use at any time, under any circumstances." Explain.

(B.Com. Vikram 1978; HPU. 1989)

References

1. Statistical Methods by S.P. Gupta.

2. Theory of Statistics by V.K. Kapoor

B.COM.–II (SEMESTER-III) PAPER : B.C. 306

BUSINESS STATISTICS

LESSON NO. 1.2 AUTHOR: DR. AMANPREET KAUR

MEASURES OF CENTRAL TENDENCY

1.2.0 Objective

The objective of the chapter are as follows:

1. To understand the calculation of arithmetic mean using different methods
2. To understand the calculation of median.
3. To understand the calculation of mode.

1.2.1 INTRODUCTION:

The presentation of data through diagrams or graphs may not convey the exact picture as desired. It is also difficult to describe the impression they give. There is, therefo need for some numerical measurement which bring out the main characteristic of the data or series. As these measures are generally in the central portion or the middle of the distribution, they are known as measures of central tendency or average. An average represents a whole series. It condences a frequency distribution into single figure. Its value lies between the minimum and maximum values of the series.

An average is a quantity which is representative of a series. It helps a person to undeerstand the significance of large mass of facts. It eliminates the unwanted details and gives a concise picture of a variable which is being studied. It helps us in grasping the central idea of a series, enables us to make comparisons and makes further analysis of data possible.

An average which truly represents a series is called a 'typical average'. An average which is not representative, but has only a theroretical value is called 'descriptive average'.

An average in whatever way defined, a particular value in a series, and as such it has to be expressed in the same units as used for collection of data. If a series give the height of trees in feet, the average should also be in feet. An average of ratios or percentages should also be expressed in ratios or percentages only.

A typical average should have the following characteristics:

(i) It should be clearly defined.

(ii) It should be representative of the data as a whole.

(iii) It should be based on all items of the series.

(iv) It should not materially change; if some more items of the same group are included at random.

(v) It should be simple to calculate.

(vi) It should be capable of mathematical treatment.

The measures of central tendency are the arithmetic mean, the median and the mode.

1.2.2 ARITHMETIC MEAN

The arithmetic mean or simply the mean is the quantity obtained by dividing the sum of the value of items in a group by the number of items. The sum of all the items in a group is known as the summation or the aggregate.

1.2.3 Calculation of the arithmetic mean in the case of ungrouped data

(a) Directed Method:

Two steps are necessary:

(i) Add all items to get the aggregate.

(ii) Divide the aggregate by the number of items.

The formula for computation is:

A.M (arithmetic mean) $a = \frac{x_1 + x_2 + x_3 .... + x_n}{n} = \frac{\Sigma x_n}{n}$

Where $x_1, x_2, ... x_n$ stand for values of different variable, Σ (sigma) or summation stands for total or aggregate or summation of items, and n for total number of items.

Example 1. Calculate the arithmetic mean of the following items:

3,4,5,6,7,8,9

It may be noted that an important feature of the arithmetic mean is that the sum of positive deviations from the mean equals the sum of negative deviations. In other words, the sum of the deviations of the various values from the mean equals zero.

The formula is Σ (x-a)=0

From the example above, the deviations (item-mean) are:

(3-6) + (4-6) + (5-6) + (6-6) + (7-6) +(8-6) + (9-6)

= (-3)+ (-2) + (-1) + 0 +1+2 +3 = -3-2-1+0+1+2+3=0

(b) Short-cut Method:

The formula from the following steps:

(i) Assume any number as average.

(ii) Find the deviations (items-assumed mean).

(iii) Add the deviations.

(iv) Divide the sum of the deviations by the total number of items.

(v) Add the quotient to the assumed or arbitrary average.

The formula is: Arithmetic mean $a = + \frac{\Sigma dx}{n}$

where A=arbitrary or assumed mean,

Σ dx = sum of deviations from the arbitrary mean.

n = total number of items.

In case of the example given above, the calculation of the arithmetic mean by the method of assumed average will be:

Value of item x	Deviation from the assumed mean dx=x-(4)
3	-1
4	0
5	+1
6	+2
7	+3
8	+4
9	+5
	Σ dx=14

Total Σ dx=14 $\therefore$Mean = $4+\frac{14}{7}$ = 4+2=6

The method has a limitation that the problem of addition may become too much laborious threfore liable to error if the number of items becomes very large.

1.2.4 Calculation of the arithmetic mean for grouped data:

(a) Equal class inverval:

The following steps help in the understanding of the formula used:

(i) Write down the mid points of the various class intervals. For finding the mid points, add the upper and lower limits of the class and divide by two.

(ii) Multiply the various mid-points by their corresponding frequencies and add these products.

(iii) Divide the sum obtained by the total number of items (i.e. sum of all the frequencies)

The formula is: $$a = \frac{f_1x_1 + f_2x_2 + f_3x_3 \ldots\ldots + f_nx_n}{n} = \frac{\sum fx}{n}$$

It is assumed that all the values lying with in each class interval are distributed evenly and that is the reason for taking the mid-points of the class intervals. It is only if the class intervals are small that the formula will give exact value of arithmetic mean. If the class intervals are large, the result will not be very exact.

Example 2. The following data gives daily wages of workers in a factory. Find the average daily wages paid to the workers:

Wages (Rs.)	Number of workers
11-13	3
13-15	4
15-17	5
17-19	6
19-21	5
21-23	4
23-25	3
	30

Solution:

Table for Computing the average daily wages:

Wages (in Rupees)	Mid Values x	No. of Workers f	fx
11-13	12	3	36
13-15	14	4	56
15-17	16	5	80
17-19	18	6	108
19-21	20	5	100
21-23	22	4	88
23-25	24	3	72
	Total	n=30	$\sum$ fx=540

$$a = \frac{\sum fx}{n} = \frac{540}{30} = \text{Rs. } 18 \text{ Ans.}$$

1.2.5 Arithmetic mean of grouped data by the short-cut method:

The steps involved are:

(i) Write down the mid-points of the various class intervals.

(ii) Assume a number as an average.

(iii) Find the deviation of mid-points from the assumed mean.

(iv) Multiply the deviations with their corresponding frequencies.

(v) Add the products obtained from (iv).

(vi) Divide the total by the total average. Thus the arithmetic mean is obtained by the formula:

$$\text{Arithmetic mean, } a = A + \frac{\sum fdx}{n}$$

Where A is assumed average, $\sum$ fdx is the sum of the products of the deviations of mid-points of classes from the assumed mean with the corresponding frequencies and n is the total number of items or sum of all the frequencies.

Carrying on with the example above, let the assumed average be 16.

Mid values M	No. of Labourers f	Deviations from the assumed average 16 (dx)	Total Deviations (fdx)
12	3	-4	-12
14	4	-2	-8
16	5	0	0
18	6	+2	+12
20	5	+4	+20
22	4	+6	+24
24	3	+8	+24
	Σ f = 30		Σ fdx=60

$$\therefore a = 16 + \frac{60}{30} = 16 + 2 = Rs.18 Ans.$$

(b) Unequal Class intervals:

A number of economic series is characterized by frequency distributions with unequal class intervals. For such frequency distributions, it is the direct method which is used for finding the arithmetic mean i.e.

$$a = \frac{\Sigma fx}{n}$$

The short-cut method described above is not so useful in computing the arithmetic mean in such cases.

Advantages:

The following are the main advantages of the arithmetic mean:

(i) It is clearly defined and is definite.

(ii) Its calculation is easy to understand.

(iii) It takes use of all the data in the series.

(iv) It enables algebraic treatment.

(v) It can be calculated from the data incomplete for other purposes e.g. if the size of the items is unknown but their aggregate and numbers are known, then the arithmetic mean can be found.

(vi) If the average of sub groups are known, average for the combined group can also be calculated.

Disadvantages:

The arithmetic mean suffers from certain draw-backs or demerits. It is not useful in all circumstances. In certain cases, it may be some what misleading. Its draw- backs or limitations are:

(i) It is considerably affected by extreme values. Any abnormal item largely affects this average. Suppose a firm pays its executive Rs. 1000 and employee three labourers whom it pays Rs. 100,50,50. The average pay will be.

$$Rs. = \frac{1000 + 100 + 50 + 50}{4} == \frac{1200}{4} \quad \text{(by adding all of them)}$$

The average pay is Rs. 300 where s 75% of the employees are really low paid.

(iii) It may not be an actual item. For example, the average of 7,11 & 12 is 10 which is different from given items. This may not be fully representative.

(iv) It cannot be correctly calculated even if a single item is neglected.

(v) The average of percentages may give fallacious results, unless actual figures are given suppose two examiners get marks as follows:

		A	B
I	House Examination	35%	65%
II	House Examination	45%	45%
III	Annual Examination	65%	35%

1.2.6 Mode

The term mode comes from the French usage 'la mode' which means the "fashion". It is the most common item of a series or data. It represents that value of a series which occurs most frequently. It is the position of the "greatest density" or the "thickest value" i.e. it is the size of the item in a series which possesses the maximum frequency. Clement Burton defined mode as the "predominant item in a group which is situated at the "average income", the reference is to a mode or a norm. If it is said that the most common income in a particular country is Rs. 5000 per annum, it means that largest number of person get Rs. 5000 per year.

1.2.7 Methods of Computation of mode:

Where the data is ungrouped, mode can be found by identifying the value which occurs most frequently. The data arranged in the form of a table. When a certain value occurs more frequently than any other value, the distribution is called unimodal. If two different values have equal and maximum frequencies associated with item, the distribution is known as bimodal. When the values in a series do not repeat themselves, mode cannot be established.

When grouping is possible or the series are either discrete or continuous the calculation of the mode can be done by applying either the grouping method or the formula method.

(i) The grouping method: In this method the frequencies against each value are written down. These frequencies are then added in two's and the total's are written in lines between the values added. The frequencies can be added in two ways:

(a) By adding frequencies of numbers 1 and 2, 3 and 4, 5 and 6, and so on.

(b) By adding frequencies of numbers 2 and 3, 4 and 5, 6 and 7 and so on.

Items can be grouped adn regrouped in three's or four's also. The procedure for grouping may be summarised as follow:

(i) Write the actual frequencies before their respective size.

(ii) Add the frequencies in two's.

(iii) Add again in two's leaving out the first frequency.

(iv) Add in three's.

(v) Add in three's leaving out the first frequency.

(vi) Add in three's leaving out the first two frequencies.

Example 3.

Find out the mode of the following series:

Size	Frequency	Arranging Size	Frequency
5	48	13	52
6	52	14	41
7	56	15	57
8	60	16	63
9	63	17	52
10	57	18	48
11	55	19	40
12	5	-	-

Solution

Since there are two values (i.e. 9 and 16) having the highest frequency, the model size can be found by the method of grouping as shown below:

Size of item	1	2	3	4	5	6
5	48					
		100				
6	52		108	156		
7	56				168	
		116				
8	60					$\boxed{179}$
			$\boxed{123}$	$\boxed{180}$		
9	$\boxed{63}$					
		$\boxed{120}$				
10	57		112		$\boxed{175}$	
						162
11	55					
		105				
12	50		102	157		
					143	
13	52					
		93				150
14	41					
			98			
15	57			161		
		$\boxed{120}$				
16	$\boxed{63}$		115		172	
						163
17	52					
		100				
18	48					
			88	140		
19	40					

Analysis Table

Columns No.	Size of item containing maximum frequency					
	7	8	9	10	11	12
1			9			10
2			6	10	15	10
3		8	9			
4		8	9	10		
5			9	10	11	
6	7	8	9			
No. of items	1	3	6	3	2	2

From the analysis table it is clear that mode is 9.

II. The Formula Method: For grouped data mode is calculated by identifying the class with the highest frequency a follows:

Mode or MO = $L_1 + \frac{D_1}{D_1 + D_2} \times h$

Where L1 = the lower limit of the model class i.e. the class with the maximum frequency:

D1 = the difference between the frequency of the modal class and that of the preceding lower class.

D2 = the different between the frequency of the modal class and that of the class following the modal class.

H = the different between the lower and upper limts of the modal class.

Example 4.

Find mode for the following data relating to income of workers in a city.

Income per year	No. of families
300-600	500
600-1200	1500
1200-2400	3000
2400-3600	6500
3600-4800	3500
4800-8000	1800
8000-15000	720
15000-above	80

$$MO = L_1 + \frac{D_1}{D_1 + D_2} \times h$$

Here the modal clas is Rs. 2400-3600 and therefore

L_1= 2400, D_1=6500-3000=3500, D_2=6500-3600=3000 and

h=3600-2400=1200

Mode i.e. $MO = 2400 + \frac{3500}{3500 \equiv 3000} \times ?1200$

The modal worker family's income per annum came to Rs. 3046.15.

Some characteristics of mode:

(i) The value of the mode is not affected by the extreme values of the series.

(ii) The algebraic manipulation of mode is not possible.

(iii) It occupies a place in the series where there is largest cases.

1.2.8 Locating Mode Graphically

In a frequency distribution the value of mode can also be determined graphically. The steps in calculation are:

(i) Draw a histogram of the given data.

(ii) Draw two lines diagonally inside the modal class bar, starting from each upper corner of the bar to the upper corner of the adjecent bar.

(iii) Draw a perpendicular line from the intersection of the two diagonal lines to the X-axis which gives us the modal value.

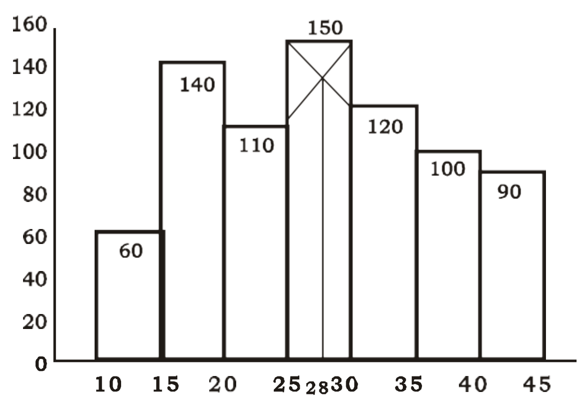
Example:

Draw a histogram for the following distribution and find the modal wage and check the value by direct calculation.

Wages in Rs.	No. of Workers
10-15	60
15-20	140
20-25	110
25-30	150
30-35	120
35-40	100
40-50	90

Solution:

The histogram of the data is given below:



It is clear from the histogram that the modal value is 28.

Merits of Mode

(i) It is very easy to locate. In a large number of cases, it can be obtained by inspection alone.

(ii) It is not influenced by the presence of a small number of extreme items.

(iii) It may be computed even if the details of extreme items are not available.

(iv) It maybe determined with considerable accuracy from well selected sample data.

Demerits:

(i) It is frequently ill-defined. In many cases, no single, well defined mode is actually found.

(ii) It is not exactly determined and differs according to the size of the sample, size of class interval, etc. Though approximate mode is easy to locate, the determination of the true mode requires extended calculation.

(iii) It is unsuitable for algebraic manipulation.

(iv) The representatives of the mode is always doubtful as it may be determined by a small number of items of similar magnitude.

(v) The product of number of items and the mode does not give the exact total of items.

(vi) It cannot be calculated by any simple arithmetic process.

Inspite of its drawbacks, the mode is increasing everyday in business and commerce.

1.2.9 Median

Median means situated in the middle. In statistics, it is the value of that item in data or series which divides the series into two equal parts, one part consisting of all values which are less and the othe of all values which are greater that median. Median is then that value of the variable which divides the total frequency into two values. It is the middle item of series which is arranged in either ascending or descending order of magnitude.

Median has the following characteristics:

(i) It is an average of position.

(ii) It depends on the number of items and not on the extreme values.

(iii) It is most typical where the central values of the series are closely grouped.

(iv) The sum of the deviations about the median if signs are ignored, will be less than the total of deviations about any other point.

Where individual values cannot be quantitatively measured e.g. in case of intelligence, health, efficiency, honesty etc., Median is useful for comparison of data.

It also help in measuring social problems like wages, distribution of income, etc. It is especially useful in cases where arithmetic mean cannot be found.

Method of calculation of the Median:

(i) Computation of Median in ungrouped Data:

To find median the data has first to be arranged in ascending or descending order of magnitude. If the number of items in a series is odd, the middle item is found by calculating the $\frac{(n+1)^{th}}{2}$ item.

Where n is number of items.

If, however, the number of items (n) is even the middle item will be in fraction and deemed to exist in the middle of the two central items, i.e. it will be the arithmetic mean of the size of $\frac{n}{2}$ and $\frac{n+2}{2}$ th items.

Example 5.

A group of 11 students obtained the following marks in English and Economics. Find out the mode of the following series:

S. No. of Student	Marks in English	Marks in Economics
1	22	46
2	29	36
3	65	30
4	33	38
5	45	39
6	50	64
7	72	50
8	33	15
9	42	42
10	25	10
11	28	72

Compare the level of knowledge of the students in the two subjects.

Solution :

The marks obtained by the students in the two subjects should first of all be arranged in an ascending order. Median marks in these subjects can then be computed.

S. No. of Student	Marks in English	Marks in Economics
1	22	10
2	25	15
3	29	30
4	33	36
5	33	38
6	42	39
7	45	42
8	48	46
9	51	50
10	65	64
11	72	72

Obviously the median in the two subject will be the size of $\frac{n+1}{2}$ th or 6th item.

$\therefore$ Median marks obtained in English = 42

Median marks obtained in Economics = 39

Hence average knowledge of students is higher in English than in Economics.

Example

Suppose instead of 11 students there are 12 students and the 12th students obtained 36 and 35 marks in the two subjects respectively. Then arranging marks in the ascending order.

No. of Student	Marks in English	Marks in Economics
1	22	10
2	25	15
3	29	30
4	33	35
5	33	36
6	36	38
7	42	39
8	45	42
9	48	46
10	51	50
11	65	64
12	72	72

Here $\frac{n}{2}$th item is 6th student and $\frac{n+2}{2}$th item is the 7th student. Therefore median will be the arithmetic mean of the score of the 6th and 7th student.

Median in English = $\frac{36+42}{2} = \frac{78}{2} = 39$

and Median in Economics = $\frac{38+39}{2} = \frac{77}{2} = 38.5$

Knowledge is a little higher in English than in Economics.

(ii) Computation of Median in a discrete series:

(a) Take down the cumulative frequencies by adding the previous frequencies.

(b) Find the mid-number i.e. Median=size of $\frac{n}{2}$th item.

(c) Find the group in which the mid-item lies.

(d) Find size of the group in which the mid-item lies.

This will be the median.

Example 6.

Find the median of the data given below:

1 Size of Item	2 Frequency	3 Size of Item	4 Frequency
4	2	11	15
5	5	12	11
6	8	13	18
7	9	14	9
8	12	15	7
9	14	16	4
10	14	17	3

Solution:

Size of Item	f	cf	Size of Item	f	cf
4	2	2	11	15	79
5	5	7	12	11	90
6	8	15	13	13	103
7	9	24	14	9	112
8	12	36	15	7	110
9	14	50	16	4	123
10	14	64	17	3	126

Median = Size of $\left(\frac{n}{2}\right)$ th item, where n=126.

$= \left(\frac{126}{2}\right) = \frac{126}{2}$ or 63rd item.

New look at c.f. column and

Median = 10

(iii) Computing of Median in a Continuous series:

The following formula is used

$$Md = L_1 + \frac{L_2 - L_1}{f}\left(\frac{n}{2} - c.f.\right)$$

where Md = median

L_1 = lower limit of the median group.

L_2 = upper limit of the median group.

F = frequency of the median group.

n is the total of frequencies.

c.f. = Cumulative frequency of the group preceding the median group.

Example 7.

Find the median of the following distribution:

Class Interval	Frequency
1-3	6
3-5	53
5-7	85
7-9	56
9-11	21

Solution:

Class Interval	Frequency	Cumulative Frequencies
1-3	6	6
3-5	53	59
5-7	85	144
7-9	56	200
9-11	21	221

Median = Value of $\frac{n}{2}$, i.e. $\frac{221}{2} = 110.5$, which lies in 5-7 group. Applying the formula:

$$Md = L_1 + \frac{L_2 - L_1}{f}\left(\frac{n}{2} - c.f.\right)$$

$$= 5 + \frac{7-5}{85}(110.5 - 59.0)$$

$$= 5 + \frac{2}{85} \; ? \; 51.5 = 5+1.21 = 6.21 \text{ Ans.}$$

1.2.10 POSITIONAL VALUES QUARTILES, DECILES AND PERCENTILES

The median divides series in two equal parts. The values of items in one part is more than the median value and the value of items in the other, less than the value of the median, with a view to have a better study about the composition of a series. It may be necessary to divide it into four, six, seven, eight, nine, ten or hundred parts.

Quartiles divide the series into 4 equal parts and there are three quartiles i.e. Q_1, Q_2, Q_3 where Q_2 is the median. Deciles divide the series into 10 equal parts and there are 9 deciles i.e. D_1, D_2,...D_9. Percentiles divide the series into 100 equal parts and there are 99 percentiles denoted as P_1, P_2,...P_{99}.

Example 7.

Class Interval	Frequency
1-3	6
3-5	53
5-7	85
7-9	56
9-11	21

Solution:

Class Interval	Frequency	Cumulative Frequencies
1-3	6	6
3-5	53	59
5-7	85	144
7-9	56	200
9-11	21	221

Q_1 is the Value of $\frac{n}{2}$, i.e. $\frac{221}{4}$ or 55.25th items which lies in 3-5 group.

Applying the formula: $Q_1 = \frac{L_2 - L_1}{4}\left(\frac{n}{4} - c.f.\right)$

$$= 3 + \frac{2}{53}\left(\frac{221}{4} - 6\right) = 3 + \frac{2}{53}(49.25) = 4.86 \quad \text{Ans.}$$

Similarly Q_3, is the size of $\frac{3n}{4}$ th item which lies in 7-9 group.

$$\therefore Q_3 = L + \frac{L_2 - L_1}{f}\left(\frac{3n}{4} - c.f.\right)$$

$$= 7 + \frac{2}{56}\left(\frac{3 \quad 221}{4} - 144\right) = 7.78 \ \textit{Ans.}$$

Merits of Median

(i) If the number of observation is odd, it is rigidly defined.

(ii) It can be easily calculated.

(iii) It can be easily understood.

(iv) It is not affected by the values of extreme items and is thus more repredentative than the arithmetic mean.

Drawbacks of Median

(i) Median is not representative of th series when the series is irregular.

(ii) An exact total cannot be obtained by multiplying the median by the number of items.

(iii) The usefulness of median is doubtful in those cases in which it is desirable to give large weight to extreme items.

(iv) The median may not be representative if the items in a series are not situated closely together.

(v) Even when the values of extremes are not known, the average can be calculated if the number of items is known.

(vi) It can be located merely by inspection in many cases.

(vii) It gives best results in cases in which direct quantitative measure is not possible, i.e. intelligence.

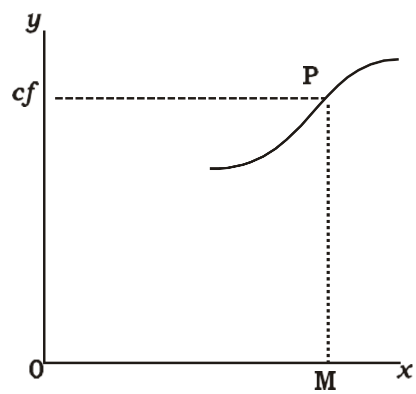
2.9.1 Uses of Median

Median is useful in the following conditions:

(i) When the values of various items cannot be mathematically measured.

(ii) When information is not available for all units.

(iii) According to W.I. King, for studies such as wages, distribution of wealth, etc. median is decidedly superior to either the mode or the arithmetic average.



Determination of Median Graphically

With the help of cumulative frequency curve median can be determined in the following way:

In the case of cumulative frequency curve (or ogive), draw a horizontal line from the mid-point of the Y-axis which represents the total frequency. This horizontal line OP meets the Ogive at P, wherefrom a perpendicular is drawn on the X-axis. This point M gives the value of the median. If two ogives (i.e. less than type and more than type) are drawn on the same graph paper, and from the point of intersection of the two curves a perpendicular is drawn on the X-axis, then that point is the median.

Example:

Draw a cumulative frequency curve from the frequency distribution and determine the median graphically.

X	1-9	9-17	17-25	25-33	23-41	41-49	49-57	57-65
f	20	30	26	16	10	8	4	4

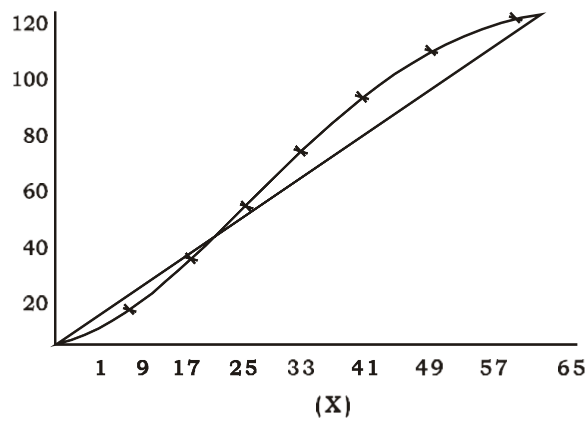
Solution:

The required curve is drawn below:

Relationship among Mean, Median and Mode

In an asymmeterical distribution

Mode=3 Median-2 Mean



$$= \frac{L_2 - L_1}{4}\left(\frac{n}{4} - c.f.\right)$$

$$= 3 + \frac{2}{53}\left(\frac{221}{4} - 6\right) = 3 + \frac{2}{53}(49.25) = 4.86$$

$$\frac{3n}{4}$$

$$\therefore Q_3 = L + \frac{L_2 - L_1}{f}\left(\frac{3n}{4} - c.f.\right)$$

$$= 7 + \frac{2}{56}\left(\frac{3 \times 221}{4} - 144\right) = 7.78 \; Ans.$$

1.2.11 Self Check Exercises

Q1: What do you mean by median?
Q2: What do you mean by mode?

1.2.12 Summary

The chapter explains the calculation of arithmetic mean by using different methods such as direct method, short cut method, step deviation method. This chapter also explains the meaning of mode and median, mode which represents the value of a series which occurs most frequently and median means value of that item in data or series which divides the series into two equal parts, one part consisting of all values which are less and the other of all values which are greater than median.

1.2.13 Glossary

AM: Arithmetic Mean

Median: It is value of that item in data or series which divides the series into two equal parts, one part consisting of all values which are less and the other of all values which are greater than median.

Mode: Represents the value of a series which occurs most frequently.

1.2.14 Exercise

(A) Short Questions

Q1: What are the advantages of arithmetic mean?

Q2: What are the advantages of median?

(B) Long Questions

Q1: What are the advantages and disadvantages of arithmetic mean?

Q2: What are the advantages and disadvantages of median?

Q3: What are the advantages and disadvantages of mode?

1.2.15 Suggested Reading

1. Statistical Methods by S.P. Gupta.
2. Theory of Statistics by V.K. Kapoor

LESSON NO. 1.3 AUTHOR: Ms. FANIZA SESHI

MEASURES OF DISPERSION

1.3.0 Objective

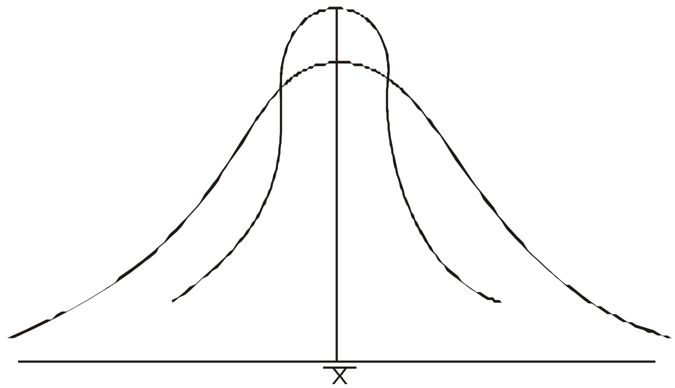
1. The main objective of this lession is to understand about the meaning of dispersion.
2. To understand about the classification of dispersion.
3. To understand the distance measure.

1.3.1 Introduction

The measures of central tendency describe that the values in the data set tend to spread (duster) around central value called average. But these values do not reveal how these values are spread (dispersed or scattered).

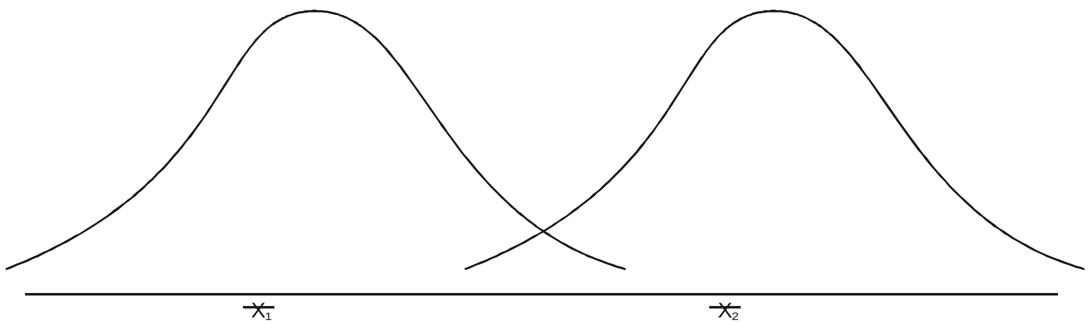
It is necessary to describe variability or dispersion of observations. Also in two or more distriobutions the average value may be same but still there can be wide disparity information of distributions. A small dispersion among values in the data set indicates that values in the data set are dustered closely around mean, Implying that mean is relaible and average. Conversely, If values in the data set are widely dustered around the mean, then this implies that mean is not reliable average, i.e. mean is not representative of data.

(a)



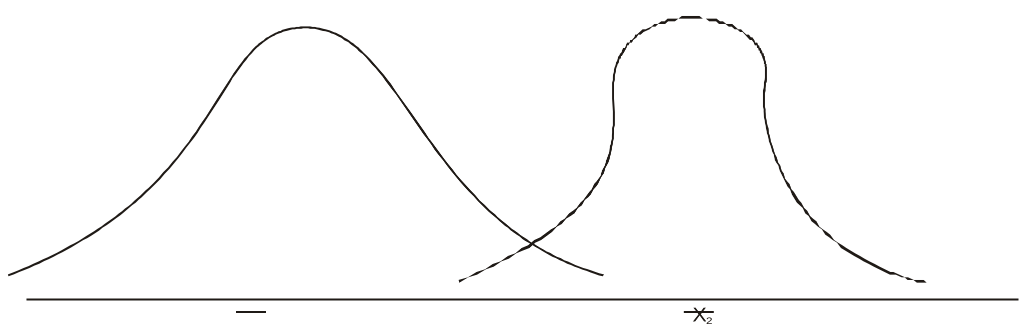
Two Distributions with same mean and different variations

(b)



Two Distributions with unequal mean and equal variations

(c)



X_1

Two distributons with unequal mean and unequal variations

Though measure of dispersion is interested in amount of degree but not the direction that is in skewness.

1.3.2 Significance of Measuring Dispersion

1. Measures of variations point out as to how for an average is representative of entire data. If variation is small than It is good estimate of the average.
2. Measurement of variations is basic to control cause of variable.
3. We can also make comparison to be made of two or more series with regard to their variability.
4. It also facilates the use of other statistical tools.

1.3.3 Classification of Measures of Dispersion

1. Absolute Measure : Variations are expressed in same statistical units in which original data are given such as Rs, kg, tonnes etc. so that we compare distributions easily.
2. Relative Measure : It is ratio of measure of absolute variation to an average. It is also called as coefficience of variation because coefficient means pure no. that is Independent of unit of Measurement.

1.3.4 Distance Measures

1. Range
2. Interquartile Deviation

1. Range

Range is difference between highest and lowest observed values. The calculation of range as a measure of dispersion is based on the location of the largest and he smallest values in the data.

Range (R) = Highest value of an observation - Lowest value of an observation.

= H–L

For example, If the smallest value of an observation in the data set is 160 and largest value is 250, then the range is 250–160 = 90

Coefficient of Range : The relative measure of range, called the coefficient of range, is obtained by applying the following formula :

$$\text{Coefficient of Range} = \frac{H-L}{H+L}$$

Example 1 : The following are the sales figures of a firm for last 12 months :

Months :	1	2	3	4	5	6	7	8	9	10	11	12
Sales : (Rs. 000)	80	82	82	84	84	86	86	88	88	90	90	92

calculate the range and coefficient of range of sales for the last 12 months.

Solution : Given that, H = 92 and L = 80, Therefore,

Range = H – L = 92 – 80 = Rs. 12

Coefficient of Range = $\frac{H-L}{H+L} = \frac{92-80}{92+80} = \frac{12}{172} = 0.069$

Advantages

1. This method is simples and easy to understand.

Disadvantages

1. It is not based on each and every observation of the distribution.
2. It cannot be computed in case of open ended distribution.

2. Interquartile Range or Deviation

As range measures based on two extreme values, to overcome this problem we have interquartile range. In this we take on quartile of observations at lower and another quartile of observations at upper end.

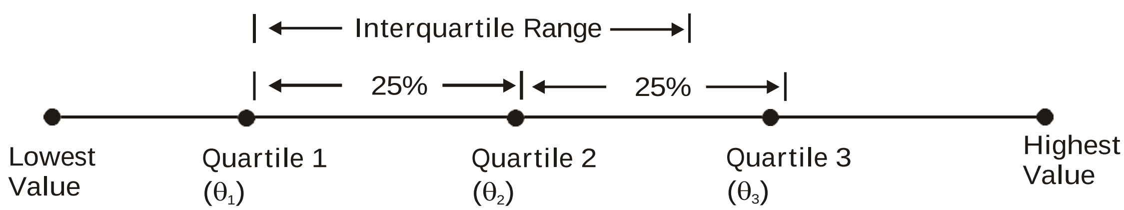
Interquartile range = $\theta_3 - \theta_1$

and quartile deviation is semi quartile range, i.e.

$$QD = \frac{\theta_3 - \theta_1}{2}$$

where; coefficient of quartile deviation is calculated as :

$$\text{Coefficient of } Q.D = \frac{\theta_3 - \theta_1}{\theta_3 + \theta_1}$$



The quartile deviation measures the average range of 25% of the values in the data set. It is computed by taking an average of the middle 50% of observed values rather than 25% part of the values in the data set.

Advantages

1. It has special utility in measuring variations incase of open and distributions.
2. It is useful in highly skewed distributions.

Disadvantages

1. It Ignores 50% items i.e. the first 25% and last 25%.

2. It does not skew scatterness around mean. It is not itself measured from an average but It is a positional average. It shows distance on scale.

1.3.5 Average Deviation Measures

Since range and interquartise deviation do not indicate how values in a data set are scattered around a central value or disperse throughout the range, therefore. It is important to measure the amount by which these values in a data set deviate from measure of central value usually mean or median. These are :

(i) Mean Absolute Deivation

(ii) Variance and standard deviation

Mean Absolute Deviation

The absolute difference between a value x_i of an observation from A.M. (or median) is called the mean absolute deviation.

In general, the mean absolute deviation is given by

$$\text{MAD} = \frac{1}{N}\sum_{i=1}^{N}|x-\mu|, \text{ for population}$$

$$\text{MAD} = \frac{1}{x}\sum_{i=1}^{N}|x-\bar{X}|, \text{ for sample}$$

For grouped frequency distribution,

$$\text{MAD} = \frac{\sum_{i=1} fi|x_i - \bar{X}|}{\sum f_i}$$

Coefficient of MAD : The relative measure of MAD is called as coefficient of MAD and is obtained by the following way :

$$\text{Coefficient of MAD} = \frac{\text{Mean absolute deviation}}{\bar{X}}$$

Example 2 : Calculate the mean absolute deviation and is coefficient from the median for the following data.

Year	Sales (Thousand)	
	Produce A	Produce B
2010	23	36
2011	41	39
2012	29	36
2013	53	31
2014	38	47

Solution : The median sales of the two products A & B is 38 and 36, respectively. The calculation in MAD in both the cases are ;

Calculation of MAD

Product A		Product B	
Sales	|X–Me| = |X–38|	sales	|X–Me| = |X–36|
23	15	31	5
29	9	36	0
38	0	36	3
41	3	39	3
53	15	47	11
x = 5	Σ|X-Me| = 42	x = 5	Σ|X–Me| = 19

Product A : $MAD = \frac{1}{x}\sum |x - Me| = \frac{42}{5} = 8.4$

$$\text{coefficient of } MAD = \frac{MAD}{Me} = \frac{8.4}{38} = 0.221$$

Product B : $MAD = \frac{1}{x}\sum |x - Me| = \frac{19}{5} = 3.8$

$$\text{coefficient of } MAD = \frac{MAD}{Me} = \frac{3.8}{36} = 0.106$$

Advantages

1. It is simple and easy to understand.
2. It is based on each and every observation of data.
3. Deviation is taken from central value.

Disadvantages

1. It does not gives an accurate Result; as It gives result when deviations taken from median. Median itself is not satisfactory measure.

Standard Deviations

Standard deviation concept introduced by Karl Pearson in 1893, widely used measure of studying the variation. It measures how much "spead" or "variability" is present in the sample, If all no. of sample are very close to zero and If all the number in sample is very close to eash other. If the numbers are dispressed then the standard deviation tends to be large. It is also known as root mean square deviations.

$$\sigma = \sqrt{\frac{\Sigma\left(X - \overrightarrow{X}\right)^2}{N}}$$

If we square the deviation, we get variation;

variance = σ^2

$\sigma = \sqrt{\text{variance}}$

σ = standard deviation

(i) A small deviation means a high degree of uniformity of observations as well as homogenity of series.

(ii) Standard deviation is useful in judging representative of mean as it is computed from mean.

Example 3 : The wholesale prices of commodity for seven consecutive days in a month are as follows :

Days	:	1	2	3	4	5	6	7
Commodity price/quintal	:	240	260	270	245	255	286	264

Calculate the variance and standard deviation from actual arithmetic mean, $\bar{x}$.

Observation (x)	$x - \bar{x} = x - 260$	$(x - \bar{x})^2$
240	–20	400
260	0	0
270	10	100
245	–15	225
255	–5	25
286	26	676
264	4	16
1820		1442

$$\bar{x} = \frac{1}{x}\Sigma x = \frac{1}{7}(1820) = 260$$

$$\text{variance } \sigma^2 = \frac{1}{x}\Sigma\,(x - \bar{x})^2 = \frac{1}{7}(1442) = 206$$

standard deviation $\sigma = \sqrt{\sigma^2} = \sqrt{206} = 14.352$

Advantages

1. It is considered as Best Measure of Variation
2. Based on every item of Distribution
3. We can also compare the variability of two or more distributions.

Disadvantages

It give more weight to extreme values and less weight to those near to mean.

1.3.6 summary

A statistical metric known as a measure of dispersion, sometimes referred to as a measure of variability, assesses the spread or variability of data values in a dataset. It offers details on how the various data points are distributed around measures of central tendency, such as the mean, median, or mode. Here is a list of several popular dispersion measures:The selection of the most appropriate measure will rely on the nature of the data and the particular topics being addressed in the analysis. Each of these measures has advantages and limits. To fully comprehend the spread of the data, it is always crucial to take into account a variety of metrics of dispersion.

1.3.7 Glossary

1. Range: The simplest measure of dispersion, it is the difference between the maximum and minimum values in the dataset. It provides a rough estimate of the spread but is sensitive to outliers.
2. Interquartile Range (IQR): The IQR is the range of the middle 50% of the data, specifically the difference between the third quartile (Q3) and the first quartile (Q1). It is less affected by extreme values compared to the range.

1.3.8 self check exercise

1. Explain variance?
2. Explain standard variance?

1.3.9 short and long question exercise

1. Write a short note on
   a. Measure of dispersion
   b. Range ?

1.3.10 long question

a. What is measure of dispersion? Explain it with the help of example ?
b. Explain average deviation measure method with the help of example?

1.3.11 suggestion reading

References :

1. Statistical Methods by S.P. Gupta.
2. Theory of Statistics by V.K. Kapoor.

LESSON NO. : 1.4 AUTHOR : DR. PARMOD K. AGGARWAL

TIME SERIES ANALYSIS - I

1.4.0 Objective

The main objective of this chapter are as follows:

1. To understand the meaning and uses of the time series.
2. To understand the components of time series.
3. To understand the methods of measuring trend.

1.4.1 Introduction

One of the most important tasks before economics and business; these days is to make estimates for the future. For example, a business man is interested in finding out his likely sales in the coming years so that he could adjust his production accordingly and avoid the possibility of either unsold stocks or inadequate production to meet the demand. However, the first step in making estimates for the future consists of gathering information from the past. In this connection one usually deals with statistical data which are collected, observed or recorded at successive intervals of time. Such data are generally referred to as 'time -series'. Thus the time series is an arrangement of statistical data in accordance with the time of its occurrence. Hence, in the analysis of time series, time is the most important, it can be week, day, hour or even minutes or seconds. Mathmetically

$y = f(t)$ where y = variable like production, population etc.
and t = time like months, years census etc.

1.4.2 Definition

- "A series is a set of statistical observation arranged in chronological order" - Morris Hamburg.
- "A time series consists of statistical data which are collected, recorded or observed over successive increments." - Patterson.

It is clear from above definitions that time series consists of data arranged chronologically. Thus if we record, the data relating to population, per capita income, prices, output, etc. for the last 5, 10, 15, 20 years or some other period, the series so emerging would be called time series.

1.4.3 Uses of the Analysis of Time Series

1. It helps in understanding the past behaviour of a variable and ill

determining the rate of growth and the extent and direction of periodic fluctuation.

2. It helps us to predict future tendencies.
3. It helps us to iron out intra-year variation. Thus seasonal ups and downs in sales may be reduced by making effective advertisements.
4. If facilitates comparison. Different time series are often compared and important conclusions drawn therefrom.

1.4.4 Components of Time Series

To classify the fluctuations of a time series, four basic types of variations are as follows :

1. Secular Trend
2. Seasonal Variation
3. Cyclical Variation
4. Irregular Variation

1. Secular Trend - T

The term 'trend' is very commonly used in day-to-day parlance. For example, we often talk of rising trend to population, prices, etc. The concept of trend does not include short-range oscillations but rather steady movements over a long time. Secular trend movements are attributable to factors such as population change, technological progress and large scale shifts in consumer tests. The underlying factor causing an upward trend in a time series may be application of natural science in the fields of agriculture and industry, the changes in the forms of business organisation facilitating accumulation of huge capital for specialisation and quality control etc. to raise the standard of living, productivity, etc.

No all time series show an upward trend. A declining rate is noticed in the data of epidemics, deaths and births, etc. owing to better and widely available medical facilities. This tendency secular movement meaning of trend :

1. The longer the period covered, the more significant the trend. To compute the trend the period must cover atleast two or three complete cycles.
2. It is not necessary that the rise or fall must continue in the same direction throughout the period.

Secular trend is usually of two types :

(a) Linear Trend : When long term rise or fall in a time series takes place by a constant amount, then that is called a linear trend. This is also known as straight line trend. This is represented by the following equation:

$$\boxed{Y = a + bX}$$

(b) Parabolic Trend : The trend is said to be parabolic when long-term rise or fall in a time-series is not taking place at a definite rate. It has many forms but most prominent of them is the second Degree Parabolic or Quadratic trend. It equation is as follows :

$$\boxed{Y = a + bX + cX^2}$$

II. Cyclical Variations - C

The term 'Cycle' refers to the recurrent variations in time series that usually last longer than a year and are regular neither in amplitude nor in length. Cyclical fluctuations are long term movements that represent consistently recurring rises and declines in activity. These movements are known as cyclical variations as they pursue an oscillating movements which, in general, takes the form of a wave, though the distance from peak to through of the waves are uneven. There are four well defined periods or phases in the business cycle, namely : (i) prosperity, (ii) decline or recession, (ill) depression, and (iv) improvement or recovery.

The study of Cyclical variations is extremely useful in framing suitable policies, for stabilizing the level of business activity.

III. Seasonal Variations - S

Seasonal variations are those periodic movements in business which occur regularly every year and have their origin in the nature of the year itself. Climate and custom together play an important role in giving rise to seasonal movements to almost all the industries. Seasonal variation is evident when the data are recorded at weekly or monthly or quarterly intervals. Although the amplitude of Seasonal Variations may vary their period is fixed being one year. As a result, seasonal variations do not appear in series of annual figures.

IV. Irregular Variations -I

Irregular variations also called 'erratic', accidental, random, refer to such variations in business activity which do not repeat in definite pattern. In fact the irregular variation is really intended to include all types of variations other than those accounting for the trend, seasonal and cyclical movements. Irregular variations are caused by such isolated special occurrences as floods, earthquakes strikes and wars. Sudden changes in demand or very rapid technological progress may also be included in this category. By their very nature these movements are very irregular and unpredictable.

1.4.5 ANALYSIS OR DECOMPOSITION OF TIME SERIES

Time-series is composed of four components i.e. Trend (T), Seasonal variations (S), Cyclical variations (C) and Irregular variations (I). There is always some sort of relationship in these four components. To study the influence of these components on time-series, they are measured separately, which is called as Analysis or Decomposition of time-series. According to Prof. Speigel; "The analysis of time series consists in decomposition of a time series into its basic components."

Models of Analysing Time Series: Analysis of time series is based on two models:

1. Additive model
2. Multiplicative model.

(1) Additive Model : This model is based on the assumption that time series is the sum of the four components. According to the formula:

$$\boxed{O = T + S + C + I}$$

O = variable

This model treats all the constituents as residuals on the basis of which, by deducting trend from the original data, short term fluctuations can be determined. Similarly, cyclical variations and irregular variations can be determined by deducting seasonal variations from short-term variations. On the basis of additive model, the analysis of various components is illustrated as given below:

$$= S + C + I$$
$$- S = C + I$$
$$O - T - S - C = I$$

(2) Multiplicative Model : This model is based on the assumption that a time series is the product of four components. According to the formula:

$$\boxed{O = T \times S \times C \times I = TSCI}$$

Whatever the component is to be separated, works as a divisor with respect to original data (O). Analysis of different components on the basis of multiplicative model can be expressed in the following forms:

$$\frac{O}{T} = SCI$$
$$\frac{O}{T \times C} = SI$$
$$\frac{O}{T \times C \times S} = I$$

Usually, multiplicative model is most often used for time series analysis.

1.4.6 METHODS OF MEASURING TREND

The main methods of measuring trend in a time-series are as follows:

(1) Free hand Curve Method.
(2) Semi-Average Method.
(3) Moving Average Method.
(4) Least Square Method.

(1) Free hand Curve Method

This is the simplest of the trend-fitting. In this method first of all the original data of the time-series is plotted on a graph-paper. Thereafter, taking care of the fluctuations of data, a smooth curve is drawn which passes through the mid-points of the fluctuations of time-series, Infact, this curve is called as freehand trend curve. This method is also called as trend fitting by inspection. The procedure of this method can be illustrated by the following example:

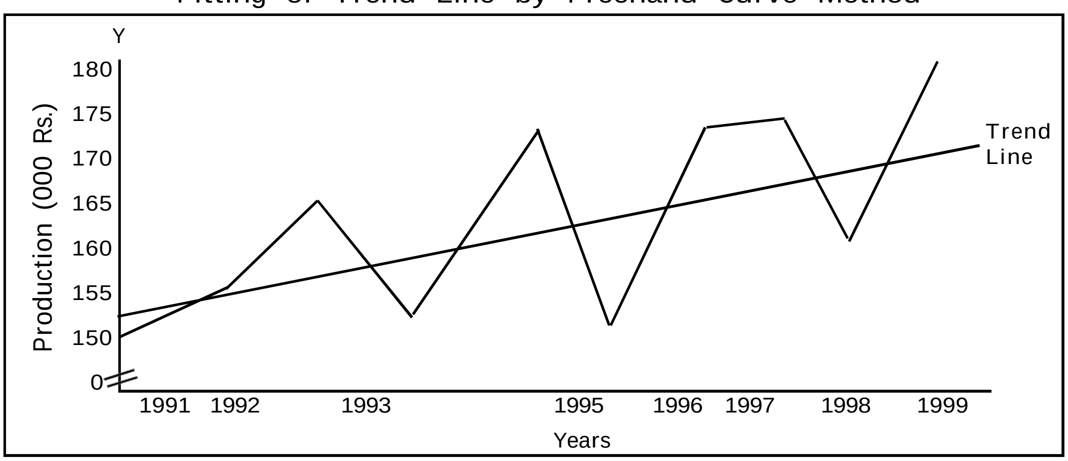
Example 1.

Fit a trend line to the following data by the free hand curve method:

Year	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Production ('000 Rs.)	150	155	165	152	174	150	174	175	160	180

Solution:

Fitting of Trend Line by Freehand Curve Method



Merits and Demerits of Freehand Curve Method

Merits

(i) This method is simple.
(ii) This method is flexible.
(iii) No mathematical formula is used in this method.
(iv) This method is also used for forecasting about future.

Demerits:

(i) This method is based on subjective judgements. So bias may affect the findings.

(ii) There is lack of accuracy in this method

(iii) Long-run movement obtained from this method is not definite. This is because a number of curves can be fitted from the same original data.

Questions for Self Practice

Fit a trend line to the following data by freehand curve method:

Year:	1986	1987	1988	1989	1990	1991	1992	1993	1994
Sales: ('0000 Units)	22	28	24	30	18	26	20	32	16

(2) Semi-Average Method

In this method, first of all time-series is divided into two equal parts and thereafter, separate arithmetic mean is calculated for each part. The trends of arithmetic means is plotted in graph corresponding to the time-periods. Joining the two points, straight line thus obtained is called as trend line. The semi-average method can be applied in case of two situations:

(1) When the number of years in a series is even: When the given number of years in a series is even like 4, 6, 8 etc., then the series can be easily divided into two equal parts. In this situation trend-fitting process can be illustrated with the following example:

Example 2.

Fit a trend line by the method of semi-average to the data given below:

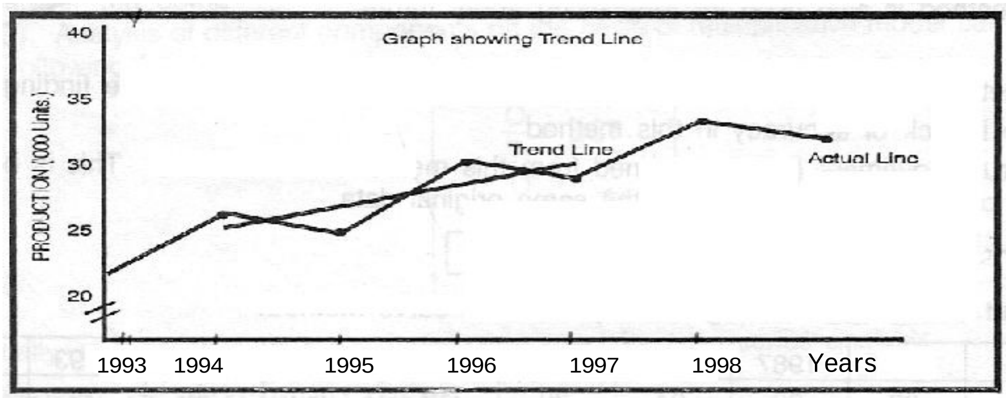
Year:	1993	1994	1995	1996	1997	1998
Production: ('000 Units)	22	26	24	30	28	32

Solution:

Fitting of Trend Line by Semi-Average Method

Year	Production ('000 units)	Semi-Total	Semi-Average	Middle Year
1993	22			
1994	26	$- \rightarrow 72$	$72 \div 3 = 24$	$- \rightarrow 1984$
1995	24			
1996	30			
1997	28	$- \rightarrow 90$	$90 \div 3 = 30$	$- \rightarrow 1987$
1998	32			

TREND LINE BY SEMI-AVERAGE METHOD



(2) When the number of years in a series is odd: When the number of years in a series is odd like 5, 7, 9, then there will be a problem in dividing the series into two equal parts. In such case, the mid-year figure is to be dropped. For example, if 1981 to 1989 (Le. 9 years) figures are given, then we will delete 1985 Le. 5th year and its corresponding figure and we will make 4-4 years' parts Le. 1981 to 1984 and 1986 to 1989. The remaining process will be the same as before. Trend fitting in this case can be illustrated by the following example:

Example 3.

Fit a trend line by the method of semi-averages to the data given below:

Year:	1991	1992	1993	1994	1995	1996	1997
Profit : ('000 Units)	20	22	27	26	30	29	40

Also estimate the profit for the year 1998.

Questions for Self Practise :

1. Fit a trend line for the following data by semi-average method:

Year:	1993	1994	1995	1996	1997	1998	
Profit : ('000 Units)	80	82	85	70	89	95	

2. Fit a trend line for the following data by semi-average method:

Year:	1990	1991	1992	1993	1994	1995	1996
Production: ('000 Units)	12	14	16	20	20	31	28

Also estimate the value for the year 1998.

(3) Moving Average Method

Under this method, moving averages are calculated. In moving average computations, one has to decide what moving year average - 3 year, 4 year, 5 year, 7 year should be taken up. The period of moving average depends upon the periodicity of data and there is no specific rule for that. The period is determined by plotting the data on the graph paper and noticing the average time interval of successive peaks or troughs. However, it is essential to consider while selecting the period of moving average that after how many years most of the fluctuations occur in the data. Moving average method is studied in two different situations:

(1) Odd Period Moving Average

(2) Even Period Moving Average.

(1) Odd Period of Moving Average : When period of the-moving average is odd, say 3 years, then following steps are to be taken for the computation of moving average:

1. First of all, add up the values corresponding to first 3 years in the time series and put the sum before the middle year (i.e. 2nd year).
2. Thereafter, leaving the first year value, add up second, third and fourth year values and put the sum in front of middle year (i.e. 3rd year). Carry this process further till we reach the last value of the series.
3. Moving totals thus obtained are to be divided by the period of the moving average and show the trend values of different years.

Similarly, five-yearly, seven-yearly moving averages can be obtained.

The computation procedure of 3-yearly moving averages can be illustrated with the following example:

Example 4.

From the following data, calculate trend values using 3-yearly moving average:

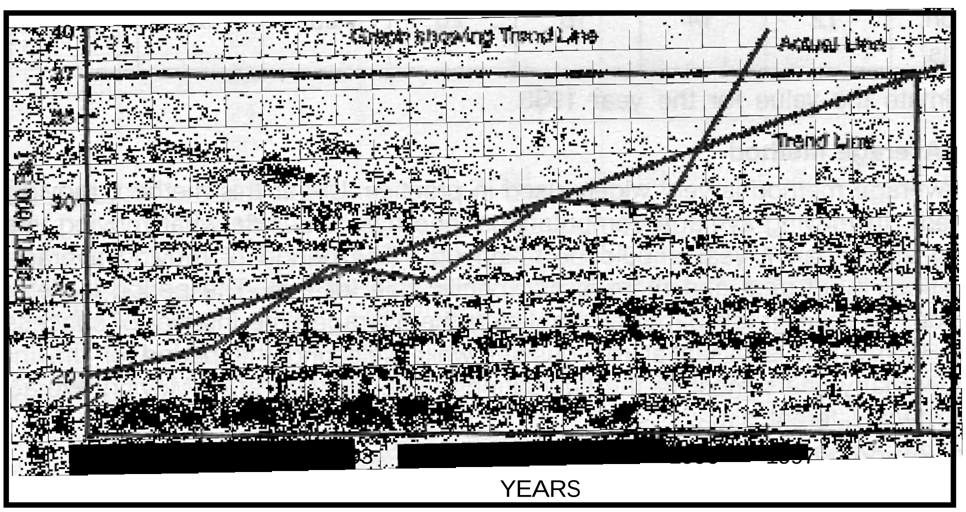
Year:	1991	1992	1993	1994	1995	1996	1997
Production:	20	22	27	26	30	29	40

Solution:

Since there are 7 years, the middle year 1984 will be left out and the arithmetic average of the two parts will be calculated as given below:

Year Year	Production ('000 units)	Semi-Total	Semi-Averge	Middle
1991	20			
1992	22	– → 69	69 ÷ 3 = 23	– → 1982
1993	27			
1994	26			Omitted
1995	30			
1996	29	– → 99	99 ÷ 3 = 33	– → 1986
1997	40			

TREND LINE BY SEMI-AVERAGE METHOD



The estimated profit for the year 1998 = Rs. 37,000.

1.4.7 Merits and Demerits of Semi-Average Method

Merits:

(i) This is an easy method.

(ii) This method is free from bias.

(iii) Trend values thus obtained are definite.

(iv) Less time and effort is involved in drawing the trend line.

Demerits:

(i) This method is based on straight line trend assumption which does not always hold true.

(ii) This method is affected by extreme values.

(iii) This method ignores the effect of cyclical fluctuations.

(2) Even Period Moving Averages: When moving average period is even, say 4 years, then moving averages have to be centered. It can be computed by two methods:

(1) First Method

(2) Second Method.

(1) First Method : The computation procedure of 4 yearly moving average is as follows:

(i) First of all, add up first 4 values corresponding to the first 4 years and put the sum in between second and third year. Thereafter the next total (i.e. from 2nd to 5th year total) is to be put in between 3rd and 4th year. Carry on this process till the last value of the series.

(ii) Now add up the 1st and 2nd 4-year totals and put them in front of 3rd year. Similarly, add up 2nd and 3rd 4-year total and put them in front of 4th year. Carryon this process till the last value.

(iii) 8 years totals thus obtained are to be divided by 8. These values are 4-yearly moving averages and show the trend values for different years.

The computation procedure is made clear by the following example:

Example 5.

Calculate the trend values using 4 yearly moving average from the following data:

Year:	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Sales: (in Crores)	7	8	9	11	10	12	8	6	5	10

Solution:

Computation of 4 Yearly Moving Average (Trend)

Year	Sales (in crore)	4 yearly moving totals	2 period moving totals of 4 yearly moving totals	4 yearly moving average (centred) (Trend values)
1990	7	-	-	-
1991	8	- $\rightarrow - 7 + 8 + 9 + 11 = 35$	-	-
1992	9		$\rightarrow - 35 + 38 = 73$	$73 \div 8 = 9.125$
		$\rightarrow - 8 + 9 + 11 + 10 = 38$		
1993	11		$\rightarrow - 38 + 42 = 80$	$80 \div 8 = 10.00$
		$\rightarrow - 9 + 11 + 10 + 12 = 42$		
1994	10		$\rightarrow - 42 + 41 = 83$	$83 \div 8 = 10.375$
		$\rightarrow - 11 + 10 + 12 + 8 = 41$		
1995	12		$\rightarrow - 41 + 36 = 77$	$77 \div 8 = 9.625$
		$\rightarrow - 10 + 12 + 8 + 6 = 36$		
1996	8		$\rightarrow - 36 + 31 = 67$	$67 \div 8 = 8.375$
		$\rightarrow - 12 + 8 + 6 + 5 = 31$		
1997	6		$\rightarrow - 31 + 29 = 60$	$60 \div 8 = 7.5$
		$\rightarrow - 8 + 6 + 5 + 10 = 29$		
1998	5	-	-	-
1999	10	-	-	-

(b) Second Method: There is an alternative method of constructing 4-yearly centered moving averages, the method of which is given below:

1. First of all add up the 4-values corresponding to the first 4 years and put the sum in between second and third year. Thereafter, the next total (i.e. from second to fifth year total) is to be put in between 3rd and 4th year. Carryon this process till the last value of the series.
2. 4-yearly moving totals thus obtained are divided by 4 to obtain 4-yearly uncentered moving averages.
3. Now add up 1st and 2nd 4-yearly moving averages and divide it by 2. Put this average in front of 3rd year. Similarly, add up 2nd and 3rd 4-yearly moving averages and divide it by 2 and put this average in front of 4th year. Carry on this process till the last value. These values so obtained are 4-yearly centered moving averages and show the trend values for different years.

1.4.8 Measurement of Short term Fluctuations

Time-series is the mixture of both trend and short term fluctuations. Therefore, if trend component is deducted from the original data, we can find out short term fluctuations.

The following example makes clear the measurement of short term fluctuations:

Example 6.

Using three yearly moving averages, determine the trend and short-term fluctuations. Plot the original and trend values on the same graph paper:

Year:	1980	1981	1982	1983	1984	1985	1986	1987
Profit: ('000 Rs.)	18	21	20	25	29	27	35	42

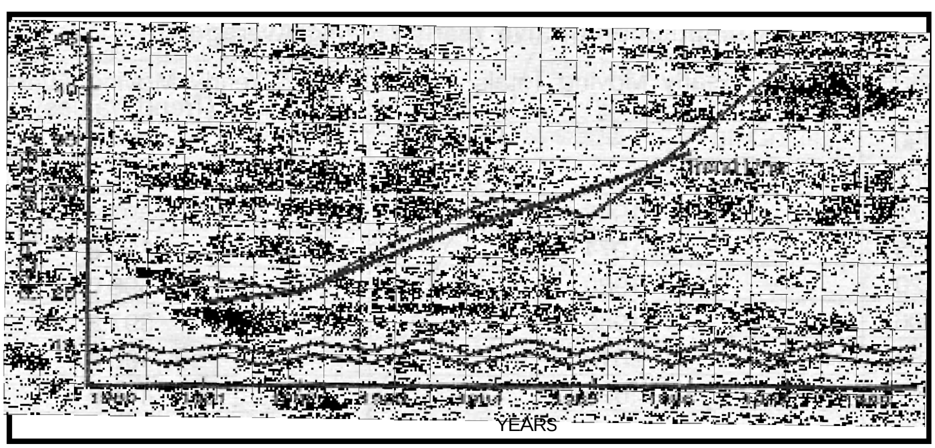
Solution:

Calculation of Trend and Short-term Fluctuations

Year	Profit (Y)	3-yearly moving totals (Trend Values- T)	3-yearly moving averages Y_c= T	Short-term Fluctuations $Y = Y_c$
1980	18	-	-	-
1981	21	59	19.667	1.333
1982	20	66	22.00	-2
1983	25	74	24.667	0.333
1984	29	81	27.000	2
1985	27	91	30.333	-3.333
1986	35	104	34.667	0.333
1987	42	-	-	-

Deducting trend values from the original series (Y), the residual left is short term fluctuations. These fluctuations have been shown in above said column (5).

GRAPH OF THE ORIGINAL AND TREND VALUES



Merit and demerits of Moving Average Method

Merits:

(1) This method is easy to understand and simple to use.

(2) This method is flexible i.e. if number of years are added in a series, previous calculations are not affected.

(3) This method is most suitable for eliminating cyclical fluctuations.

(4) This method has great practical usefulness.

Demerits :

(1) It is difficult to ascertain the proper period of moving averages and if proper period is not ascertained, results will be misleading and inaccurate.

(2) The second defect of this method is that some beginning years' and some terminal years' trend values remain beyond· the scope of calculations.

(3) The limitations of arithmetic mean affect this method adversely.

(4) If periodicity in the series is not clearly visible, this method should not be used.

Questions for Self Practice :

1. Find trend values for the following data, by using 5 yearly moving averages. Also plot the actual data and trend values on a graph :

Year	1999	1990	1991	1992	1993	1994	1995	1996	1997	1998
Production	672	679	690	702	712	802	807	809	816	821

[Ans. 691.0, 717.0, 742.6, 766.4, 789.2, 811.0]

2. Calculate trend values (using 3 yearly moving average) for the following data:

Year	1981	1982	1983	984	1985	1986	1987	1988
Production ('000)	60	72	65	80	85	85	95	110

[Ans. 69, 70.67, 73.33, 76.67, 86.67, 96.67]

3. Find trend values using 4 yearly moving average form the following data:

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Production ('000)	100	105	115	90	95	85	80	65	75	70	75	80

[Ans. 101.88, 98.75, 91.88, 84.38, 78.75, 74.38, 71.88, 73.13]

(4) Least Square Method

This is the best method of trend-fitting in a time series and is most used in practice. This is a mathematical method and a trend line in this method is fitted or obtained in such a way that following two conditions are fulfilled:

(1) $\Sigma (Y - Y_c) = 0$ i.e. the sum of the deviations of the actual values of Y and computed trend values (Y_c) is zero.

(2) $\Sigma (Y - Y_c)^2$ is least i.e. the sum of the squares of the deviations of the actual and computed trend values from this line is least.

Trend-line thus fitted under this method is called as the Line of Best Fit.

Least square method can be used to fit straight line trend or parabolic trend.

(A) Fitting of Straight Line Trend

A straight line trend can be expressed by the following equation:

$$\boxed{Y = a + bX}$$

Where Y = Trend Values, X = Unit of Time

a is the Y-intercept and b is the slope of the line.

In the above equation, to determine two constants, a and b, the following two normal equations are solved:

$$\Sigma Y = Na + b \Sigma X \qquad \text{(i)}$$

$$\Sigma XY = a \Sigma X + b \Sigma X^2 \qquad \text{(ii)}$$

After determining the equation Y = a + bX, we find the trend values related to different years and plot them on the graph paper which show a straight line trend.

There are two methods of computing straight line trend by using least square method:

(1) Direct Method

(2) Short Cut Method

(1) Direct Method

The procedure to compute straight line trend in this method is as follows:

(i) Any year other than the middle year is taken as the year of origin. Usually first year or before that is taken as zero, deviations of other years are marked on 1, 2, 3 etc. Time deviations are denoted by X:

(ii) Then ΣX, ΣY, ΣXY and ΣX^2 are computed.

(iii) The values computed are put in the following normal equations:

$$\Sigma Y = Na + b \Sigma X \qquad \text{(i)}$$

$$\Sigma XY = a \Sigma X + b \Sigma X^2 \qquad \text{(ii)}$$

The values of a and b are determined by solving the above said two normal equations.

(iv) Finally, the calculated values of a and b are put in Y = a + bX and trend values are computed. The following example makes clear the procedure of this method:

Example 7.

Fit a straight line trend by the method of least square (taking 1978 as year of origin) to the following data:

Year	1989	1990	1991	1992	1993	1994
Production (lakh tons)	5	7	9	10	12	17

Also obtain the trend values.

Solution:

Fitting of Straight Line Trend

Year	Production Y	Deviation from 1978 X	XY	X^2
1989	5	1	5	1
1990	7	2	14	4
1991	9	3	27	9
1992	10	4	40	16
1993	12	5	60	25
1994	17	6	102	36
N=6	$\Sigma Y = 60$	$\Sigma X = 21$	$\Sigma XY = 248$	$\Sigma X^2 = 91$

The straight line trend is defined by the equation:

$$Y = a + bX$$

Two normal equations are

$$\Sigma Y = Na + b \Sigma X$$

$\Sigma XY = a \Sigma X + b \Sigma X^2$ Substituting the values, we get

$$60 = 6a + 21b \qquad \text{(i)}$$

$$248 = 21a + 91b \qquad \text{(ii)}$$

Solving the two equations (i) and (ii).

Multiplying (i) by 7 and (ii) by 2 and then subtracting

$$\begin{array}{l} 420 = 42a + 147b \\ 496 = 42a + 182b \\ - \quad\ - \qquad\quad - \\ \hline -76 = \qquad -35b \\ \hline \end{array}$$

$$b = \frac{-76}{-35} = 2.17$$

By substituting the value of 'b' in equation (i), we get!

60 = 6a + 21b

or 60 = 6a + 21(2.17)

or 6a = 14.43

$\therefore$ a = 2.40

Hence, the trend equation is

Y = 2.40 + 2.17X; origin = 1978, X unit = 1 Year.

Computation of Trend Values

For 1989, X=1, Y	2.40 + 2.17 (1)	= 4.57
For 1990, X=2, Y	2.40 + 2.17 (2)	= 6.74
For 1991, X=3, Y	2.40 + 2.17 (3)	= 8.91
For 1992, X=4, Y	2.40 + 2.17 (4)	= 11.08
For 1993, X=5, Y	2.40 + 2.17 (5)	= 13.25
For 1994, X=6, Y	2.40 + 2.17 (6)	= 15.42

Questions for Self Practice

1. Fit a straight line trend by the method of least square (taking 1981 as year of origin) to the following data :

Year	1991	1992	1993	1994	1995
Value	15	21	25	33	40

[Ans. : Y = 14.4+6.2X, 14.4, 20.6, 26.8, 33, 39.2]

Also obtain the trend values.

2. Fit a straight line trend by the method of least square to the following data :

Year	1989	1990	1991	1992	1993
Value	45	56	78	46	75

[Ans. : Y = 45 + 5X; 50, 55, 60 65, 70]

Also obtain the trend values.

(Take the year 1968 as origin year).

(2) Short Cut Method

The process of computation in this method to find straight-line trend is as follows :

1. First of all, middle-year is taken as the year of origin with value zero and deviations for other years are computed. Sum of the deviations will be zero i.e. $\sum X=0$. Since deviations above middle-year will be –1, –2, –3 etc. and deviations

after middle year will be 1, 2, 3... etc. and deviations above and below middle year will balance out. This is made clear by the following example:

Year	1992	1993	1994	1995	1996	1997	1998
X:	–3	–2	–1	0	–1	+2	+3

2. ΣY, ΣXY and ΣX^2 are computed.

3. For computing the values of a, b, we need not have normal equations but they are found by the following formulae:

$$a = \frac{\Sigma Y}{N} ; b = \frac{\Sigma XY}{\Sigma X^2}$$

4. Finally, the calculated values of a, b are put in the equation Y = a + bX and with its help, trend values are computed.

Short-cut method is studied in two cases:

(i) When number of years are odd

(ii) When number of years are even.

(1) When Number of Years are Odd : When number of years is odd like 5, 7, 9 ... etc. then the computation of straight line trend can be illustrated with the following examples:

Example 8.

Fit a straight line trend by the method of least squares to the following data and also show on graph paper:

Year	1993	1994	1995	1996	1997	1998	1999
Production ('000 units)	90	90	92	83	94	99	92

Solution:

Year	Y	Deviations from 1983 X	XY	X^2
1993	80	-3	-240	9
1994	90	-2	-180	4
1995	92	-1	-92	1
1996	83	a	a	()
1997	94	1	+94	1
1998	99	2	+198	4
1999	92	3	+276	9
N=7	ΣY=630	ΣX=0	ΣXY = 56	ΣX^2=28

Let the equation of the straight line trend is:

$Y = a + bX$

Since $\sum X = 0$

$\therefore \quad a = \frac{\sum Y}{N}$

$b = \frac{\sum XY}{\sum X^2}$

Substituting the values, we get

$a = \frac{\sum Y}{N} = \frac{630}{7} = 90$

$b = \frac{\sum XY}{\sum X^2} = \frac{56}{28} = 2$

Thus, $Y = 90 + 2X$

Origin = 1996, X unit = One Year.

Computation of Trend Values

For 1993, $X = -3$, $Y_c = 90 + 2\,(-3) = 84$

For 1994, $X = -2$, $Y_c = 90 + 2\,(-2) = 86$

For 1995, $X = -1$, $Y_c = 90 + 2\,(-1) = 88$

For 1996, $X = 0$, $Y_c = 90 + 2\,(0) = 90$

For 1997. $X = +1$, $Y_c = 90 + 2\,(1) = 92$

For 1998, $X = +2$, $Y_c = 90 + 2\,(2) = 94$

For 1999, $X = +3$, $Y_c = 90 + 2\,(3) = 96$

(2) When Number of Years is Even: When given number of years is even (6, 10, 12 ... etc.), in such a case, what is to be the middle year becomes a problem. In such a case, mean of the two middle years is taken as year of origin i.e. zero value and correspondingly, deviations for other years are found out. Deviations will be -2.5, -1.5, -0.5, 0.5, 1.5 and 2.5. To simplify the computation process, these deviations are multiplied by 2. The remaining steps are the same as before.

Example 9.

Fit a straight line trend by the method of least square to the following data and find the trend values i.e. at years :

Year	1991	1992	1993	1994	1995	1996	1997	1998
Value :	80	90	92	83	94	99	92	104

Solution:

Fitting of Straight Line Trend

Year	Value Y	Deviations from 1964.5	Deviations Multiplied by 2 X	XY	X^2
1991	80	-3.5	-7	-560	49
1992	90	-2.5	-5	-450	25
1993	92	-1.5	-3	-276	9
1994	83	-0.5	-1	-83	1
1995	94	+0.5	1	94	1
1996	99	+ 1.5	3	297	9
1997	92	+2.5	5	460	25
1998	104	+3.5	7	728	49
N=8	$\Sigma Y = 734$		$\Sigma X=0$	$\Sigma XY=210$	$\Sigma X^2 = 168$

The equation of the straight trend line is:

$$Y = a + bX$$

Since $\Sigma X = 0$

$$\therefore \quad a = \frac{\Sigma Y}{N}, \qquad b = \frac{\Sigma XY}{\Sigma X^2}$$

Substituting the values, we get

$$a = \frac{734}{8} = 91.75$$

$$b = \frac{210}{168} = 1.25$$

The straight line trend is:

$Y = 91.75 + 1.25X$; Origin 1964.5 X unit $= \frac{1}{2}$ year

Computation of Trend Values:'

For 1991, $X = -7,\ Y_c = 91.75 + 1.25\ (-7)$ $= 83$
For 1992, $X = -5,\ Y_c = 91.75 + 1.25\ (-5)$ $= 85.5$
For 1993, $X = -3,\ Y_c = 91.75 + 1.25\ (-3)$ $= 88$
For 1994, $X = -1,\ Y_c = 91.75 + 1.25\ (-1)$ $= 90.5$
For 1995, $X = +1, Y_c = 91.75 + 1.25\ (+1)$ $= 93.0$
For 1996, $X = +3, Y_c = 91.75 + 1.25\ (+3)$ $= 95.5$
For 1997, $X = +5, Y_c = 91.75 + 1.25\ (+5)$ $= 98.5$
For 1998, $X = +7, Y_c = 91.75 + 1.25\ (+7)$ $= 100.5$

Aliter : Instead of calculating the trend values like this, we can double the value of b (since b gives half yearly trend values) and add to the preceding trend value.

Annual trend value = b = 1.25 ×2 = 2.50

Y_{1992} = 83 + 2.5 = 85.5

Y_{1993} = 85.5 + 2.5 = 88 and so on.

Merits and Demerits of Least Square Method

Merits :

(1) This method is far better than moving average method because the trend values for all the years are obtained. Not even a single initial or terminal trend values is left over in this method.

(2) It results in a mathematical equation which may be used for forecasting.

(3) It is widely used method of fitting a curve to the given data. The results obtained are reliable and appropriate.

Demerits :

(1) The computation process in this method is complex which is not easily understandable.

(2) This method does not have the attribute of flexibility. If some figures are added to or subtracted from the original data, all computations have to be redone.

(3) It is difficult to select an appropriate type of equation in this method. Results based on inappropriate selection of equation are likely to be misleading.

Exercise

1. Fit for straight line trend by method of least squares to the following data and show on graph paper :

Year	1981	1982	1983	1984	1985	1986	1987
Production ('000 tons)	80	90	92	83	94	99	92

[Ans. Y = 89 + 2X, 84, 86, 88, 90, 92, 94, 96]

2. Fit for straight line trend by method of least squares to the following data and also tabulate the trend values :

Year	1971	1972	1973	1994	1975	1976	1977
Value :	77	80	94	85	91	98	90

[Ans. Y = 89 + 2X; 85; 87; 89; 91; 93; 95]

3. Fit a straight line trend by method of least squares and estimate the production

for the year 1995 and 1997:

Year	1989	1990	1991	1992	1993	1994
Production (in lakh tons)	25	40	47	59	68	80

[Ans. Y = 53.17 + 5.3X, 90.27, 111.47]

4. Fit a linear trend to the following time series by the method of least squares and also obtain the trend values:

Year	1954	1955	1956	1957	1958	1959
Production (in crores of Rs.)	7	10	12	14	17	24

[Ans. Y = 14 + 1.54X, 6.28, 9.37, 12.46, 15.55, 18.64, 21.73]

1.4.8 Self check exercise

Q1: Write down the components of time series?

Q2: What do you mean by linear trend?

1.4.9 Summary

This chapter explain the meaning of time series which means a time series consists of statistical data which are arranged chronologically. The time series can be used in understanding the past behavior of the variable and also help in predicting future tendencies, facilitates comparison. There are various components of time series which are explained in the chapter such as Secular Trend, Seasonal Variation, Cyclical Variation, Irregular Variation. This chapter also explain the method of measuring trend analysis and their merits.

1.4.10 Glossary

1. Time Series Data: Time series consists of statistical data which are arranged chronologically.
2. Parabolic Trend: It is parabolic when long term rise or fall in a time series is not taking place at a definite rate.

1.4.11 Exercise

(A) Short Questions

Q1: What do you mean by time series data?

Q2: What are the uses of analysis of time series data?

(B) Long Questions

Q1: Explain the components of time series?

Q2: Explain the model of analyzing time series?

Q3: Explain the method of measuring trend?

1.4.12 References :

1. Statistical Methods by S.P. Gupta.
2. Theory of Statistics by V.K. Kapoor.

LESSON NO. 1.5 AUTHOR : DR. PARMOD K. AGGARWAL

TIME SERIES ANALYSIS - II

1.5.0 Objective

1.The main objective of this lession is to understand the measurement of time series
2.To understand these method with the help of exemples.

1.5.1 INTRODUCTION

In economics and trend analysis, the knowledge of seasonal variations is also very useful. With its help he can, on the one hand, make short-term planning for his business activities and on the other hand, he can immune himself from the effects of short-period variations. Therefore, analysis of seasonal variations is very important. In this chapter, we will discuss the methods of measuring seasonal variations.

1.5.2 MEASUREMENT OF SEASONAL VARIATIONS

The main methods of measuring seasonal variations are as follows:

(1) Method of Simple Averages
(2) Method of Moving Average
(3) Ratio to Moving Average
(4) Ratio to Trend Method
(5) Link Relatives Method

Let us consider them in detail.

(1) Method of Simple Averages

This is the simplest method of measuring seasonal variations. This method is used in those situations where trend is assumed to be absent in the data. This method involves the following steps:

(1) The given data is arranged monthwise or quarterwise for different years.
(2) The totals of each month or quarter for different years are obtained and then dividing the sum by 12 or 4, the average of each month or quarter is computed.
(3) The average of the monthly average or quarterly average is then computed.
(4) Taking the general average as base, seasonal indices for each month or quarter are computed by using the following formulae:

(i) When monthly data is given

General Average

$$\text{Seasonal Index for Jan.} = \frac{\text{Average of Jan.}}{\text{General Average}} \times 100$$

Seasonal Index for Feb. = $\frac{\text{Average of Feb.}}{\text{General Average}} \times 100$

Similarly, seasonal indices for other months can also be computed:

(ii) When quarterly data is given

Seasonal Index for I Quarter = $\frac{\text{Average of I Quarter}}{\text{General Average}} \times 100$

Seasonal Index for II Quarter = $\frac{\text{Average of II Quarter}}{\text{General Average}} \times 100$

Similarly, seasonal indices for III and IV quarters can also be computed.

The following example illustrates the procedure of this method:

Example 1.

Assuming that trend is absent, determine if there is any seasonality in the data given below:

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
2001	37	41	33	35
2002	37	39	36	36
2003	40	43	33	31

What are the seasonal indices for various quarters?

Solution:

Computation of Seasonal Indices

Quarters	2001	2002	2003	Quarterly totals for 3 years	Quarterly Average
Q_1	37	37	40	114	38
Q_2	41	39	43	123	41
Q_3	33	36	33	102	34
Q_4	35	36	31	102	34

Average of Average = $\frac{38+41+34+34}{4} = \frac{147}{4} = 36.75$

(or General Average)

Seasonal Index = $\frac{\text{Quarterly Average}}{\text{General Average}} \times 100$

Seasonal Index for 1st Quarter = $\frac{38}{36.75} \times 100 = 103.40$

$$\text{Seasonal Index for 2nd Quarter} = \frac{41}{36.75} \times 100 = 111.56$$

$$\text{Seasonal Index for 3rd Quarter} = \frac{34}{36.75} \times 100 = 92.52$$

$$\text{Seasonal Index for 4th Quarter} = \frac{34}{36.75} \times 100 = 92.52$$

1.5.3 QUESTION FOR SELF PRACTISE

1. Calculate the seasonal index for the following data by using simple average method for the following data:

Year	Summer	Monsoon	Autumun	Winter
2001	112	110	120	115
2002	80	145	105	90
2003	95	100	140	80
2004	110	90	130	100

[Ans: 92.2, 103.39, 114.5, 89.4]

(2) Method of Moving Average

This method is superior than the simple average method. It involves the following steps:

(1) First of all, moving averages are computed from the original data. If data are given on monthly basis, then 12-monthly moving-averages and if quarterly data are given, then 4-quarterly moving averages are computed.

(2) After this, from each original figure, corresponding trend value is deducted to arrive at short term fluctuations.

(3) By making a separate table, short term fluctuations for different months or quarterly periods are summed up and the sum is divided by the number of years. These arithmetic averages are the seasonal indices for different months or quarters.

This method can be illustrated by the following example:

Example 2.

Find out the seasonal fluctuations by the method of moving average from the following data:

Year	Summer	Monsoon	Autumn	Winter
2001	30	81	62	119
2002	33	104	86	171
2003	42	153	99	221
2004	56	172	129	235
2005	67	201	136	302

Solution

Years	Values (0)	4-Quarterly Totals	2 Period Totals Centralized Totals	Average (T) Trend	Quarterly Short-term (O- T)	Seasonal (S)
2001 S. I	30	-	-	-	-	-
M. II	81	-	-	-	-	-
		- → 292				
A. III	62		- → 587	73	-11	-19
		- → 295				
W. IV	119		- → 613	77	+42	+68
		- → 318				
2002 S. I	33		- → 660	83	-50	-75
		- → 342				
M. II	104		- → 736	92	+12	+25
		- → 394				
A. III	86		- → 797	100	-14	-19
		- → 403				
W. IV	171		- → 855	107	+64	+68
		- → 452				
2002 S. I	42		- → 917	115	-73	-75
		- → 465				
M. II	153		- → 980	123	+30	+25
		- → 515				
A. III	99		- → 1044	131	-32	-19
		- → 529				
W. IV	221		- → 1077	135	+86	+68
		- → 548				
2004 S. I	56		- → 1126	141	-85	-75
		- → 578				
M. II	172		- → 1170	146	+26	+25
		- → 592				
A. III	129		- → 1195	149	-20	-19
		- → 603				
W. IV	235		- → 1235	154	+81	+68
		- → 632				
2005 S. I	67		- → 1271	159	-92	-75
		- → 639				
M. II	201		- → 1345	168	+33	+25
		- → 706				
A. III	136	-	-	-	-	-
W. IV	302	-	-	-	-	-

Calculation of Seasonal Variations
(from short-term fluctuations)

Year	Summer	Monsoon	Autumun	Winter
2001	-	-	-11	+42
2002	-50	+12	-14	+64
2003	-73	+30	-32	+86
2004	-85	+26	-20	+81
2005	-92	+33	-	-
Total	-300	101	-77	+273
Average	-75	+25 app.	-19 app.	+68 app.

QUESTION FOR SELF PRACTISE

1. Find the seasonal fluctuations by the method of moving averages from the following data:

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1991	74	76 -	74	80
1992	82	68	50	62
1992	70	74	70	82

[Ans.: 6.25, 1.375, -8.50, 0.875]

(3) Ratio to Moving Average Method

This is the most popular method of measuring seasonal variations. The following steps are taken up under this method:

1. Obtain the trend values by the method of moving average method. If given data are quarterly, then 4-quarterly moving averages are found out. As against it, if given data are monthly, then 12-monthly moving averages are computed.
2. After this, each figure relating to the time-periods of original data is divided by the corresponding trend value and the quotient is multiplied by 100 to get ratio-to-moving average:

$$\text{Ratio to Moving Average} = \frac{O}{T} \times 100$$ Where O = Original Value

T = Moving Average

3. Next, arithmetic averages are computed after arranging the ratio-to-moving averages related to different periods in a separate table.
4. All averages relating to ratio-to-moving averages are summed up and treated to get a general average.

5. Finally, making the general average as base, the seasonal indices for quarters are found by the following formula:

$$\text{Seasonal Indices} = \frac{\text{Quarterly Average}}{\text{General Average}} \times 100$$

Thus method can be illustrated with the following example:

Example 3.

From the following data, calculate seasonal indices by the Ratio to Moving Average Method.

Year	I Quarter	II Quarter	III Quarter	IV Quarter
1991	68	62	61	63
1992	65	58	66	61
1993	68	63	63	67

Solution:

Calculation of Seasonal Indices by Ratio-to-Moving Average Method

Years	Quarter Period	Values (0)	4 Quarterly Moving	2 Totals	4 Quarterly Moving Average(T)	Ratio to Moving Average $\left(\frac{O}{T}\times 100\right)$
1991	I	68	-	-	-	-
	II	62	-	-	-	-
			→ 254			
	III	61		→ 505	63.1	96.7
			→ 251			
	IV	63		→ 498	62.3	101.1
			→ 247			
1992	I	65		→ 499	62.4	104.2
			→ 252			
	II	58		→ 502	62.8	92.4
			→ 250			
	III	66		→ 503	62.9	104.9
			→ 253			
	IV	61		→ 511	63.9	95.5
			→ 258			
1993		68		→ 513	64.1	106.1
			→ 255			
	I	63		→ 516	64.5	97.7
			→ 261			
	III	63	-	-	-	-
	IV	67	-	-	-	-

Calculation of Seasonal Indices

Year	I Quarter	II Quarter	III Quarter	IV Quarter
1991	-	-	96.7	101.1
1992	104.2	92.4	104.9	95.5
1993	106.1	97.7	-	-
Totals	210.3	190.1	201.6	196.6
Average	105.15	95.05	100.8	98.3
Seasonal Indices	105.4	95.2	101.0	98.5

General Average = $\frac{105.15+95.05+100.8+98.3}{4} = \frac{399.3}{4} = 99.8$

Calculation of Seasonal Indices

Seasonal Indices for I Quarter = $\frac{105.15}{99.8} \times 100 = 105.36$

II Quarter = $\frac{95.05}{99.8} \times 100 = 95.24$

III Quarter = $\frac{100.8}{99.8} \times 100 = 101.00$

IV Quarter = $\frac{98.3}{99.8} \times 100 = 98.5$

QUESTION FOR SELF-PRACTICE

1. Calculate the seasonal index for the data given below by Ratio to Moving Average Method.

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1980	68	62	61 63	
1981	65	58	56 61	
1982	68	63	63 67	
1983	70	59	56 62	
1984	60	55	51 58	

(4) Ratio-to Trend Method

Under this method, the following steps are taken up for the measurement of seasonal variations:

1. Obtain the trend values season-wise (quarterly) by the method of least squares.

2. By dividing the each value of original data (0) relating to all the periods by the corresponding trend value (T), ratio-to-trend is computed. Symbolically

$$\text{Ratio to Trend} = \frac{O}{T} \times 100$$

3. The arithmetic mean of each quarterly or monthly period ratio-to-trend is computed.
4. General Average is computed by summing up all the averages relating to the quarters or months. Hereafter, seasonal indices are computed by using the following formula:

$$\text{Seasonal Index} = \frac{\text{Quarterly Average}}{\text{General Average}} \times 100$$

This method can be illustrated with the following example:

Example 4.

Find out seasonal Index by ratio-to-trend method from the data given below:

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1991	30	40	36 34	
1992	34	52	50 44	
1993	40	58	54 48	
1994	54	76	68 62	
1995	80	92	86 82	

Solution:

First we will determine the trend values for yearly data by fitting a straight line trend by the method of least squares.

Year	Yearly Total	Quarterly Average	X Deviations from 1989	XY	X^2	Average Quarterly Trend values (Yc)
1991	140	35	-2	-70	4	32
1992	180	45	-1	-45	1	44
1993	200	50	0	0	0	56
1994	260	65	+1	+65	1	68
1995	340	85	+2	+170	4	80
N=5		$\Sigma Y = 280$	$\Sigma X=0$	$\Sigma XY = 120$	$\Sigma X^2 = 10$	

The equation of the straight line trend is

$$Y = a + bX$$

Since $\Sigma X = 0, a = \frac{\Sigma Y}{N}, b = \frac{\Sigma XY}{\Sigma X^2}$

$\therefore \quad a = \frac{280}{5} = 56$

$b = \frac{120}{10} = 12$

$\therefore \quad Y = 56 + 12 X$

Annual Increment = b = 12

Quarterly Increment $\frac{12}{4} = 3$

For 1991, X = -2, Y = 56 + 12 (-2) = 56 - 24 = 32

Other trend values can be found by adding the value of b in the preceding trend values.

1.5.4 Calculation of Quarterly Trend Values

Consider 1991 Trend value for the middle quarter i.e. half of the 2nd and half of the 3rd is 32. Quarterly increment is 3. So the trend value of 2nd quarter is 32 - 3/2 i.e. 30.5 and for 3rd quarter is 32 + 3/2 i.e. 33.5. Trend value for the 1st quarter is 30.5 - 3 i.e. 27.5 and of 4th quarter is 33.5 + 3 i.e. 36.5. Thus, we get quarterly trend values which are given below:

Quarterly Trend Values

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1991	27.5	30.5	33.5	36.5
1992	39.5	42.5	45.5	48.5
1993	51.5	54.5	57.5	60.5
1994	63.5	66.5	69.5	72.5
1995	75.5	78.5	81.5	84.5

Quarterly Values as % of Trend Values

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1991	109.1	131.1	107.5	93.1
1992	86.1	122.4	109.5	90.7
1993	77.7	106.4	93.7	79.3
1994	85.0	114.3	97.8	85.5
1995	106.0	117.1	105.5	97.0
Total	463.9	591.3	514.6	445.6
Average	92.78	118.26	102.92	89.12
Seasonal Indices	$\frac{92.78}{100.77}\times 100$ = 92.0	$\frac{118.26}{100.77}\times 100$ = 117.4	$\frac{102.92}{100.77}\times 100$ = 102.1	$\frac{89.12}{100.77}\times 100$ = 88.4

$$\text{General Average} = \frac{92.78+118.26+102.92+89.12}{4} = \frac{403.8}{4} = 100.77$$

QUESTIONS FOR SELF PRACTISE

Using 'Ratio to Trend' method, determine the quarterly seasonal indices for the following data:

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
2001	60	80	72 68	
2002	68	104	100	88
2003	80	116	108	96
2004	108	152	136	124
2005	160	184	172	164

[Ans.: 92.03, 117.33, 102.10, 88.44]

(5) Link Relatives Method

This is the most difficult method of measuring the seasonal variations. The steps involved in this method are as follows:

1. Calculate the link relatives of the seasonal figures - monthly or quarterly. For this, the following formula is used:

$$\text{Link Relative} = \frac{\text{Current season's figure}}{\text{Previous season's figure}} \times 100$$

2. Then the average of the link relatives for each month or quarter is computed.

3. The average of the link relatives are then converted into chain relatives. For this, the following formula is used:

   Chain Relatives =

$$\frac{\text{Av. of LR of the current season's figure} \times \text{Chain Relatives of the previous season's figure}}{100}$$

4. The chain relative for the 1st term is calculated on the basis of current relatives of the last term. For this, the following formula is used:

   Chain Relatives of the 1st term =

$$\frac{\text{Chain Relatives of the last seasonal's figure} \times \text{Av. of LR of the 1st season}}{100}$$

5. Theoretically, chain relative of the first period should be 100 but sometime, due to the influence of the trend, this can be more than or less than 100. The difference in this case be found out by deducting 100 from the revised chain relative of the first term. This difference is divided by the number of periods and the quotient is multiplied by 1, 2, 3, etc. Values thus obtained are, subtracted from the chain relative of 2nd term, chain relative of the 3rd term and the chain relative of the 4th term.

6. Finally, arithmetic mean of the adjusted or corrected chain relatives is computed. By taking general average as base, seasonal indices are computed by the following formula:

$$\text{Seasonal Indices} = \frac{\text{Season's Average}}{\text{General Average}} \times 100$$

This method is illustrated by the following example:

Example 5.

Calculate seasonal indices from the following data by using link relatives method:

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
2001	20	40	60	80
2002	30	30	40	90
2003	40	60	30	120
2004	50	50	70	150

Solution:

Calculation of Seasonal Indices by Link Relative Method

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
2001	-	200	150	133.3
2002	37.5	100	133.3	225
2003	44.4	150	50	400
2004	41.7	100	140	214.3
Total of L.R.	123.6	550	473.3	972.6
Average of L.R.	41.2	137.5	118.3	243.2
Chain Relatives	100	$\frac{100 \times 137.5}{100} = 137.5$	$\frac{137.5 \times 118.3}{100} = 162.7$	$\frac{162.7 \times 243.2}{100} = 395.7$
Corrected Chain Relatives	100	137.5–15.75 = 121.75	162.7–31.5 = 131.2	395.7–47.25 = 348.45
Seasonal Indices	$\frac{100 \times 100}{175.35} = 57.03$	$\frac{121.75 \times 100}{175.35} = 69.43$	$\frac{131.2 \times 100}{175.35} = 74.82$	$\frac{348.45 \times 100}{175.35} = 198.72$

Chain relative of the first quarter = $\frac{41.2 \times 395.7}{100} = 163$

(on the basis of the last quarter)

The difference between these chain relatives = 163 - 100 = 63

Difference per quarter = $\frac{63}{4}$ = 15.75

Adjusted (or corrected) chain relatives are obtained by subtracting 1 × 15.75, 2 × 15.75, 3 × 15.75 from the chain relatives of the 2nd, 3rd and 4th quarters respectively.

Average of corrected chain relative = $\frac{100 + 121.75 + 131.2 + 348.45}{4} = 175.35$

(or General Average)

Seasonal variation indices have been calculated as follows :

Seasonal variation indices = $\frac{\text{Corrected Chain Relative}}{\text{General Average}} \times 100$

QUESTION FOR SELF PRACTISE

1. Calculate seasonal indices by link relative method from the following data :

Link Relatives

Quarter	1991	1992	1993	1994	1995
I	-	80	88	80	83
II	120	117	129	125	117
III	133	113	111	115	120
IV	83	89	93	96	79

1.5.4 summary

Time series data is a sequence of data points collected over successive time intervals. It is a common data type in various fields, including finance, economics, weather forecasting, and many others. Analyzing time series data allows us to identify patterns, trends, and seasonal fluctuations to make predictions and informed decisions. Time series analysis is a powerful tool for understanding historical data, identifying patterns, and making predictions, contributing to decision-making and planning in various domains.

1.5.5 Glossary

1. Evaluation Metrics:

To assess the accuracy of time series forecasts, various evaluation metrics are used, such as mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), or mean absolute percentage error (MAPE).

2. Moving Average (MA):

The Moving Average is a method used to smooth out short-term fluctuations and highlight long-term trends in a time series. It involves calculating the average of a fixed number of consecutive data points (the "window") and placing the average at the center of the window.

1.5.6 self check erercise

1. Explain time series?
2. Explain ratio to trend method?

1.5.7 Short question

1. Write a short note on seasonal index?
2. Explain advantages of time series?

1.5.8 Long question Exercise

2. What is a seasonal index? Explain the different methods of estimating it.
3. Discuss the ratio-to moving average and the ratio-to trend method of measuring seasonal variations. Compare the two methods.
4. Explain any method of estimating the seasonal index for a time series based on quarterly data.
5. Describe, step by step, the moving average method of determining seasonal index.
6. Explain briefly the various methods of isolating seasonal fluctuations in time series.
7. Apply method of Link Relatives in order to Calculate seasonal Indices :

Quarter	1992	1993	1994	1995	1996
I	6.0	5.4	6.8	7.2	6.6
II	6.5	7.9	6.6	5.8	7.4
III	7.8	8.4	9.3	7.5	8.0
IV	8.7	7.3	6.4	8.5	7.1

[Ans. 88.09, 94.44, 113.05, 104.43]

References :

1. Statistical Methods by S.P. Gupta.
2. Theory of Statistics by V.K. Kapoor.

Mandatory Student Feedback Form

https://forms.gle/KS5CLhvpwrpgjwN98

Note: Students, kindly click this google form link, and fill this feedback form once.

Mandatory Student Feedback Form

https://forms.gle/KS5CLhvpwrpgjwN98

Note: Students, kindly click this google form link, and fill this feedback form once.