# BUSINESS STATISTICS

## B.Com. SECOND YEAR

### Semester – 3

**Lesson Writers**

**Dr. R. Jaya Prakash Reddy**,
                    M.Com., Ph.D.
Reader in Commerce,
SGHR & MCMR College,
Guntur.

**Dr. K. Kanaka Durga,**
                    M.Com., M. Phil., Ph.D.
Reader in Commerce,
H.O.D, Dept. of Commerce,
Hindu College, Guntur.

**Dr. T. Nageswara Rao**
                    M.Com., Ph.D.
Principal (Retd.),
Hindu College, Guntur.

**Dr. V. Naga Nirmala**,
                    MBA, M.Com., MIB, M.Phil., Ph.D.
Faculty,
Acharya Nagarjuna University.

### Editor

**Prof. V. Chandra Sekhara Rao**
M.Com., Ph.D.
Dept. of Commerce & Management Studies,
Acharya Nagarjuna University.

### Director

**Dr. NAGARAJU BATTU**
MBA, MHRM, LLM, M.Sc.(Psy), M.A.(Soc), M.Ed., M.Phil., Ph.D.
Centre for Distance Education,
Acharya Nagarjuna University,
Nagarjuna Nagar 522 510, GUNTUR.

**Ph : 0863 - 2346208, 2346222, 2346259 ( Study Material)**
**Website : www.anucde.info**
**e-mail : anucdedirector@gmail.com**

**B.Com . SECOND YEAR :  Semester - 3**

# BUSINESS STAISTICS

**First Edition   :  2023**

**No. of Copies  :**

This book is exclusively prepared for the use of students of B.Com. programme, Centre for Distance Education, Acharya Nagarjuna University and this book is meant for limited circulation only.

**Printed at  :**

# FOREWORD

Since its establishment in 1976, Acharya Nagarjuna University has been forging ahead in the path of progress and dynamism, offering a variety of courses and research contributions. I am extremely happy that by gaining a 'A' Grade from the NAAC in the year 2014, the Acharya Nagarjuna University is offering educational opportunities at the UG, PG levels apart from research degrees to students from over 285 affiliated colleges spread over the two districts of Guntur and Prakasam.

The University has also started the Centre for Distance Education with the aim to bring higher education within reach of all. The centre will be a great help to those who cannot join in colleges, those who cannot afford the exorbitant fees as regular students, and even housewives desirous of pursuing higher studies. With the goal of bringing education in the door step of all such people. Acharya Nagarjuna University has started offering B.A, and B, Com courses at the Degree level and M.A, M.Com., L.L.M., courses at the PG level from the academic year 2021-22 on the basis of Semester system.

To facilitate easier understanding by students studying through the distance mode, these self-instruction materials have been prepared by eminent and experienced teachers. The lessons have been drafted with great care and expertise in the stipulated time by these teachers. Constructive ideas and scholarly suggestions are welcome from students and teachers invited respectively. Such ideas will be incorporated for the greater efficacy of this distance mode of education. For clarification of doubts and feedback, weekly classes and contact classes will be arranged at the UG and PG levels respectively.

It is aim that students getting higher education through the Centre for Distance Education should improve their qualification, have better employment opportunities and in turn facilitate the country's progress. It is my fond desire that in the years to come, the Centre for Distance Education will go from strength to strength in the form of new courses and by catering to larger number of people. My congratulations to all the Directors, Coordinators, Editors and Lesson -writers of the Centre who have helped in these endeavours.

**Prof. P.Rajasekhar**
**Vice –Chancellor,**
**Acharya Nagarjuna University**

## PROGRAMME : THREE – YEAR  B.Com
## (General & Computer Applications)

### Domain Subject : Commerce
Semester – wise Syllabus under CBCS

(w.e.f. 2020 – 21 Admitted Batch)

### II Year B.Com(Gen. & CA) : Semester – III
### 307 BCO 21 – Course 3B : Business Statistics

**Learning Outcomes :**

At the end of the course the student will be able to :

- Understand the importance of Statistics.
- Formulate complete, Concise and correct mathematical proofs.
- Frame problems using multiple mathematical and statistical tools measuring relationships by standard techniques.
- Build and assess data – based models.
- Learn and apply the statistical tools in day to day life.
- Create quantitative models to solve real world problems in appropriate contexts.

## Syllabus :

**Unit – 1 : Introduction to Statistics :**

Definition – Importance – Characteristic and Limitations of Statistics – Classification and Tabulation – Frequency Distribution Table – Diagrams and Graphic Presentation of Data (including problems).

**Unit – 2 : Measures of Central Tendency :**

Types of Averages  – Qualities of Good – Mean, Median, Mode and Median based Averages Geometric Mean – Harmonic Mean (including problems).

**Unit – 3 : Measures of Dispersion :**

Measures and Properties of Dispersion – Absolute and Relative – Types of Dispersion – Range – Quartile Deviation (Semi – Inter Quartile Range) – Mean Deviation – Standard Deviation – Coefficient of Variation (including problems).

**Unit – 4 : Skewness and Kurtosis :**

Measures of Skewness : Absolute and Relative Measures – Coefficient of Skewness : Karl Pearson's and Bowley's and Kelly's – Kurtosis :  Meso kurtosis, Platy kurtosis and Lepto kurtosis (including problems).

**Unit – 5 : Measures of Relation  :**

Meaning and use of Correlation – Types of  Correlation – Karl Pearson's Correlation Coefficient – Probable Error – Spearman's Rank Correlation (including problems).

**References :**

1.    Business Statistics , Reddy C.R., Deep  Publications.

2.    Statistical  Methods  : Gupta S.P., Sultan Chand & Sons.

3.    Statistical – Problems and Solutions : Kapoor V.K. Sultan Chand & Sons.

4.    Fundamentals of Statistics : Elhance D.N.

5.    Business Statistics , Dr. P.R. Vittal Margham Publications.

6.    Business Statistics , L.S. Agarwal, Kalyani Publications.

7.    Statistics : Dr.V. Murali Krishna Seven Hills International Publishers.

8.    Fundamentals of Statistics : Gupta S.C., Sultan Chand & Sons.

9.    Statistics – Theory, Methods and Applications  : Sancheti D.C.& Kapoor V.K.

10.   Business Statistics : J.K. Sharma , Vikas Publishers.

11.   Business Statistics : Bharat Jhunjhunwala, S. Chand Publishers.

12.   Business Statistics : S.L. Aggarval, S.L.Bharadwaj and K. Raghuveer Kalyani Publishers.

**Suggested Co-Curricular Activities :**

- Student Seminars, Quiz
- Problem solving Exercises
- Observe Live Population Clocks – India and World.
- Collection of Statistical Data of Village / Town, District, State, Nation.
- Participate in Crop Cutting Experiments at Villages.
- Percentiles in CET Exams.
- Practice Statistical Functions in MS Excel.

- Draw Diagrams and Graphs in MS Excel.
- Use statistical tools in real life like class / college results, local production etc.
- Prepare questionnaire and schedule.
- Application of Averages in every day life.
- Examinations (Scheduled and Surprise tests) on all units.
- Any similar activities with imaginative thinking beyond the prescribed syllabus.

# MODEL QUESTION PAPER

## B.Com. – Second Year

## Semester - 3

## BUSINESS STATISTICS

**Time: Three hours**                                                **Max. Marks: 70**

### SECTION A             ( 5 x 4 = 20 Marks)

Answer any **FIVE** of the following questions.

1. Write any five uses of tabulation.

2. Write any five limitations of statistics.

3. Write any five objectives of graphical presentation of data.

4. Write any five qualities of a good average.

5. Write the concept of median.

6. Write any five objectives of range.

7. Write a brief note on Skewness.

8. What is Correlation?

### SECTION B             (5 x 10 = 50 Marks)

Answer the following questions.

9. (a) Explain the characteristics of statistics.

<div align="center">Or</div>

   (b) Draw a line diagram to present following data.

| Class : | M.Com. | M.Sc. | M.A. | M.B.B.S. | B.E |
|---|---|---|---|---|---|
| No. Of students : | 220 | 180 | 340 | 80 | 120 |

10. (a) Calculate the mode from the following :

| X : | 30 – 50 | 50 – 70 | 70 – 90 | 90 – 110 | 110 – 130 | 130 – 150 | 150 – 170 |
|---|---|---|---|---|---|---|---|
| F : | 7 | 17 | 34 | 48 | 32 | 14 | 3 |

(b) From the following data calculate the Arithmetic Mean

| Class Interval : | 0 – 4 | 4 – 8 | 8 – 12 | 12 – 16 | 16 – 20 | 20 – 24 |
|---|---|---|---|---|---|---|
| Frequency : | 9 | 14 | 36 | 28 | 10 | 3 |

11. (a) Compute Quartile Deviation and coefficient of Quartile Deviation for the following data.

| X : | 10 | 12 | 14 | 16 | 18 | 20 | 22 | 24 | 28 | 30 | 34 | 36 | 38 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| F : | 3 | 6 | 10 | 15 | 20 | 24 | 30 | 22 | 18 | 14 | 10 | 6 | 6 |

(b) Calculate the standard deviation for the following data :

| Sl. No. : | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| X : | 40 | 42 | 38 | 56 | 15 | 71 | 82 | 65 | 37 | 80 |

12. (a) From the following data calculate Karl Pearson's Coefficient of Skewness.

| Distance K.M : | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 |
|---|---|---|---|---|---|---|---|---|---|
| No. Of Buses : | 16 | 20 | 29 | 49 | 61 | 42 | 23 | 8 | 2 |

(b) Explain Kurtosis and give its measurements.

13. (a) Calculate Rank correlation from the following data :

| Ranks A : | 1 | 3.5 | 3.5 | 2 | 6 | 5 | 8.5 | 8.5 | 7 |
|---|---|---|---|---|---|---|---|---|---|
| Ranks B : | 2 | 4 | 3 | 1 | 5 | 6 | 7.5 | 9 | 7.5 |

(b) From the following data calculate Karl Pearson's Coefficient of correlation for the data :

| X : | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Y : | 20 | 35 | 60 | 100 | 120 | 130 |

******

# CONTENTS

## LESSON - 1
## STATISTICS - SCOPE AND IMPORTANCE

**OBJECTIVES :**

After going through the lesson you will be able to understand the following:

1. Meaning and definitions of Statistics
2. Functions of Statistics
3. Scope and importance of Statistics
4. Limitations of Statistics

**STRUCTURE OF THE LESSON :**

1.1 Meaning of Statistics

1.2 Definitions of Statistics

1.3 Characteristics of Statistics

1.4 Functions of Statistics

1.5 Scope and Importance of Statistics

1.6 Limitations of Statistics

1.7 Summary

1.8 Glossary

1.9 Self -Assessment Questions

### 1.1 MEANING OF STATISTICS :

The word statistics is generally used in two ways: one as '*data',* and the other as '*methodsin statistics'*. In the case of the first one, statistics stands for data. The statistics (data) of rice production in India is an example of this type. Such statistics are found wherever records are collected and maintained in numerical and quantitative forms. Here the use of the word 'Statistics'is in a plural sense employed to denote only a collection of facts in figures.

In the second case also the word is used in plural form. It stands for all the principles and devices used in the collection, analysis and interpretation of quantitative statements of facts.

When the word statistics is used as a science of statistics, it is used in the **singular form**, denoting just a branch of applied mathematics. It is also customary to use the word 'statistics' which stands for a measure of formula employed in statistical studies, like an average, dispersion, coefficient of correlation etc.

### 1.2 DEFINITIONS OF STATISTICS :

Statistics has been defined variously by different authors in different times. The following are some of the important definitions of Statistics.

"Science of Counting" — Bowley.

"Science of estimates and probabilities"     —   Boddington.

"Statistical methods are methods specially adapted to the elucidation of quantitative date effectedby a multiplicity of causes" — Yule.

"Statistics is the method of judging collective natural or social phenomena from the results obtainedby the analysis of an enumeration or collection of estimates" — W.I. King.

"The science of Statistics is a study of the methods applied in collecting, analyzing and interpreting quantitative data, effected by multiple causation in any department of enquiry" — Ghosh and Chaudhry.

"Classified facts respecting the condition of the people in a state …. especially those facts whichcan be stated in numbers or in tables of numbers or in any tabular or classified arrangement" —Webster.

"Statistics are numerical statements of facts in any department of enquiry placed in relation to each other" — Bowley.

"Statistics are measurements, enumerations or estimates of natural or social phenomena, systematically arranged so as to exhibit their inter-relations" — Connor.

"Statistics is concerned with scientific methods for collecting, organizing, summarizing, presenting and analyzing data, as well as drawing valid conclusions and making reasonable decisions on thebasis of such analysis" — Murray R. Siegel.

The last two definitions given above can be said as reasonably adequate definitions. Fromthe above definitions, we can understand that Statistics must possess the following characteristics:

## 1.3  CHARACTERISTICS OF STATISTICS :

**1. Numerical statements of facts :** Statistics are numerical facts. If they are described in qualitative manner they should be reduced to definite numerical quantities. For example, good, average andpoor are qualitative terms. To understand in quantitative terms, they should be defined as – goodstudents are those who secure over 60% marks, those securing between 40 and 60% are averagestudents and those below 40% are poor.

**2. Aggregates of facts :** Statistics do not take into account individual cases. One student gets first class marks or that he is a good student, does not constitute Statistics unless the total number of students appearing in the examination is given out, of which so many passed, and in such and such divisions. Studies pertaining to individuals are not significant from statistical point of view, for conclusions cannot be drawn by means of comparison and also the figure cannot be treated otherwise. In order to advance the study it is necessary that other observations must be made available.

**3. They should be capable of being related to each other :** It is not significant as to how many students have passed in an examination unless it is known how many appeared, how these figures compare with similar figures of the previous years, and how do they compare with the figures of other sections of the same class, etc.

**4. They must have certain objects behind them :** Statistics must be collected for a pre-determinedpurpose. The figures must relate to a department of enquiry. Sets of figures without any object behind them are not capable of being placed in relation to each other.  If in a school there are

500students and 15 teachers, these figures may constitute statistics, because here the object may be to find the student-teacher ratio, but if instead of teachers we give the strength of class IV employees, there is obviously no object behind such a study. All aggregates of facts must pertain to a department of inquiry in order that they may be designed as Statistics.

**5. They are affected to a marked extent by a large number of causes :** There should not be only a single factor responsible for bringing about a change in the series. As the height increases, the weight also increases. It is a physical phenomenon. But the increase in weight is not caused by an increase in height alone; there are a large number of other factors also, viz., climate, diet, racial characteristics etc. If there is only one factor operating at a time, the study ceases to be significant from statistical point of view.

**6. Reasonable standard of accuracy must be maintained in collection of statistics :** Statisticsdeal with numbers. Sometimes they have to deal with very large numbers so mush so that it becomes impossible to observe each one of the items individually. It, then, becomes necessary to observe a sample and to apply the result to the entire group. We must be satisfied if the results ofthe smaller group are almost identical to those of the larger group. The term 'reasonable standard' is relative, depending upon the object of the enquiry and the resources available.

## 1.4  FUNCTIONS OF STATISTICS :

The following are the various functions of statistics.

1.  **Measurement Phenomena :** Statistics provides measurement to social phenomena. In this respect it has two types of functions to perform. If there is already a scale of measurement we try to collect data according to it and if there is no standard scale of measurement we try to provide one through statistical analysis and evaluation of variables involved. Thus the first category of functions includes collection of all types of data. Someof the data can be collected by means of actual counting while others have to be estimated.

The other function of Statistics is providing standard scale of measurement where it doesnot exist. Most of the social phenomena are qualitative in nature and we do not have standard scale of measurement. For example, we generally say that the standard of living of a person is high or low, but we cannot give the exact measurement of it. Index numbers and scaling techniques of Statistics provide quantitative measurement.

2.  **Description of facts :** Statistics provides description of fact by means of numbers. We can know about the magnitude of child marriages, or drinking through the statistics of these facts. Similarly, we can have a clear picture of the unemployment situation in the country only when wehave the figures of unemployed people, duration of unemployment, the type of work that they cando size of their family, any supplementary source of income and so on. Statistics tries to introducefurther clarity by means of the use of graph, diagrams, charts etc.

    i.  **Objective valuation of phenomena :** Qualitative descriptions are generally subjective in natureand may differ according to persons own idea of its magnitude. This gives rise to the lack of uniformity. Statistics, by providing standard scale helps eliminating element of subjectivity. Differentpeople may give different impression regarding the crime situation in a country but when we expressit in numbers there can be only one description. Statistics thus helps in objective and accurate valuation of a social phenomenon.

    ii.  **Trends and Estimates :** Statistics tries to find out the direction and magnitude of change in a phenomenon over time. With the help of these we can find out its position in the near or distant future by projecting the trend further. For example, we

generally find that the population of a countrytends to rise regularly. By measuring the rate of growth we can forecast population on any futuredate.

    **iii.**     **Comparative study :** Statistics provides the facility of comparative analysis. This comparison may be on the basis of time, place or facts. Comparison is made possible through quantitative measurement. For example, the health of two towns can be compared through death rate. Intelligence of two or students can be compared by means of Intelligent Quotient (I.Q). By giving the figures for the crime we can compare the administrative efficiency and police administration oftwo places. Statistics by providing a common measurement helps in the comparison. The change in the price level overtime can be compared by means of index numbers.

    **iv.**     **Degree of relationship :** With the hope of statistical analysis we try to establish relationship between any two or more variables. This is done through various complicated statistical measures like coefficient of correlation, association of attributes, co-variance etc. The more important thing about statistical inference is that we not only find out that two variables are correlated but we canalso locate the degree of relationship.

## 1.5     SCOPE AND IMPORTANCE OF STATISTICS :

Statistics has become as wide as to include in its fold all quantitative studies and analysisrelating to any department of enquiry. This, indeed, give the science of Statistics a very wide scope and one would thing that Statistics has almost an unlimited scope.

The chief importance of Statistics lies in providing the quantitative measurement to a phenomenon. Lord Kelvin rightly says, "when you can measure what you are speaking about and express it in numbers, you know something about it, but when you cannot measure it, when you cannot express it in numbers your knowledge is of a meager and unsatisfactory kind". Quantitative measurement is the sign of the growth of particular discipline and our knowledge and control overthe phenomenon. We are no longer satisfied by casual remark that the prices are rising, we mustknow how much they have risen. This we can do my means of index numbers of prices. It will nolonger satisfy us to say that India has economically improved since independence. We would liketo have exact measurement. This can be provided by figures of national income and per capital income.

Quantification of social phenomena is the basis of objective observation. Qualitative description is by nature haphazard, not standard and subjective. If two persons are asked to comment about the standard of living of a person, they are very likely to give different opinions. This can be avoided only when we have found out an exact measurement of the standard of living.Similarly, if intelligence of boys were to be expressed in the qualitative terms it would not give a clear picture to us. But if the same were to be expressed in terms of examination marks or I.Q. there will be no difficulty in understanding it and also there will not be variety of opinions about it.

Statistical analysis brings greater precision to our thinking. When facts are reduced to arithmetical figures all argument comes to an end and conclusion can be challenged only by counter Statistics. Figures never lie. They will put plain facts in the coldest and most detached way whatever may be the outcome.

As we are moving more towards social planning we have to base our policy upon aggregativefigures. This is not much of consequence whether a person has committed suicide under some strange circumstances, what is important for social planners are the fact whether there has beena fall in the number of suicides. We can never remove suicides from the society. What is of consequence is

therefore, whether the number of crimes and their seriousness in increasing or decreasing. Despite of our best efforts the accidents must occur. We as social planners are mainly concerned whether the accidents have shown a declining tendency.

Statistics is equally important in the evaluation of social reforms and nature and extent of social evils. Nothing can give clearer picture of the evil of drinking than the figures regarding the cases of suicides, indebtedness, high death rate and incidence of disease in the families of those who are drunkards. Similarly, the usefulness of prohibition can also be judged by the facts.

Statistical methods are becoming more and more popular among the social sciences. Successful attempts have been made at providing standard quantitative measures of phenomena which ahs hitherto remained qualitative in nature. We are moving more towards perfection and precision with the use of these refined tools of analysis.

## 1.6 LIMITATIONS OF STATISTICS :

Like other sciences, statistics also has its limitations. They are as follows:

i. **Unable to express quantitatively :** Statistics cannot be applied to those facts which are not capable of being quantitatively expressed. Such facts should first be reduced to precise quantitative terms. For example, we cannot compare 'culture' of two countries unless we specify by 'culture' of two countries unless we specify by culture so many industries, hospitals, educational institutions, places of worship, law courts, etc. Statistical studies

cannot be brought to bear upon such phenomena unless we express them in definite mathematical quantities. Similarly, it is not possible to study 'prosperity', 'intelligence', 'honesty', 'youth' etc., unless we specify them as standing for certain requisite quantitative standards.

ii. **Not applicable to studies of individuals :** Statistics does not take cognizance of individual items because they are aggregates of facts. It is unimportant as to what are the marks secured by a student in a certain class test, unless we know the marks of all the students and draw conclusions on that basis. Marks of one student do not constitute statistics, because one of the characteristics of Statistics is that they should be capable of being placed in relation to each other. Individual items cannot be placed in such a relationship

iii. **Statistical laws are true only on dan average and in the long run :** The quantitative nature of Statistics is true only on an average and in the long run. For example, the theory of probability says that if we toss a coin twice, one time it may fall head upward and a second time head downward. But it is possible that both the times it may come head upward or head downward. This possibility of 50 per cent times heads upward and 50 per cent times head downward will be approximately true if this experiment is repeated a larger number of times.

iv. **Statistics often leads to false conclusions :** Statistics often leads to false conclusions, generally, in cases where Statistics are quoted without context or details. For example, in a certain competitive examination in the subject Computers the students of Andhra University have done better than those of Osmania University, it does not mean that the former University has a better standard. It is possible that the students of Andhra University may have been trained in special course in Computers while those of Osmania University may not have enjoyed such facility.

v. **Uniform data always not possible :** The statistical data must be uniform and its main characteristics must be stable throughout the study. It is not possible to compare the wages

in two factories if the average wage is composed of adult wages in one, and of the wages of adults and children in the other. The data must be highly uniform and homogeneous.

vi.    **Only one among various methods :** Statistical methods are not the only method of finding the value of a group. There are other methods of studying a problem besides statistics.

vii.    **Wrong handling :** Statistics must always be handled by experts; otherwise, they give wrong results.

viii.    **Distrust of Statistics :** There is a popular feeling that statistics is undesirable. According to Gladstone, 'there are three degrees of comparison in lying – lies, damned lies and Statistics'. There can be no more damaging statement than this regarding the utility and seriousness of purpose of Statistics with which skilled students of the science works.

It is however, a mistake to apply these limitations to Statistics only. There are other sciences also which suffer from these limitations. Some of these limitations emerge from the very nature of the science. For example, statistics is applicable to quantitative studies only; so also is Mathematics, Astronomy etc. Naturally, therefore, when these limitations of Statistics are described, it is often forgotten that these are the features which distinguish Statistics from other sciences. Hence these should not be stated as its limitations. Similarly, the laws of Statistics are true on an average and not necessarily in all cases, just as there are exceptions of laws and rules in other social sciences. So far as the drawing of fallacious conclusions is concerned, such wrong conclusions are capable of being drawn in all cases where precise understanding of problems is lacking. Then Statistics requires the data to be uniform. Such uniformity is necessary in all cases where comparisons are to be made. Uniformity is required to render the data comparable on an equal footing. Similarly, this should not be considered as a limitation of Statistics that there are other methods besides Statistics to study a problem, just as there are several systems of medicine by which a particular disease can be cured. We do not say that the existence of the several systems of medicine is a limitation of each one of them. Thus, Statistics have limitations like any other science, and such limitations can be avoided if it is used by experts in this field.

## 1.7   SUMMARY :

Statistics is inevitable for any type of quantitative measurement. It is characterized by numerical statements of facts, aggregate of facts etc. It functions as a measurement phenomenon; it describes the facts and it finds out the direction and magnitude of change in phenomenon over time. Statistics has a wide scope and its importance is well known in all spheres of quantitative measurement. Though it has certain limitations like any other science, they can be avoided mostly, if statistical tools and methods are used by experts in the field.

## 1.8   GLOSSARY :

**Statistics:** It is the science which deals with the collection, classification and tabulation of numerical facts as the basis of explanation, description and comparison of phenomena.

## 1.9   SELF ASSESSMENT QUESTIONS :

1. Define Statistics and explain its characteristics.

2. "Statistics is the science of counting". Give the functions of Statistics.

3. Explain the importance, scope and limitations of statistics.


**Dr. R. Jayaprakash Reddy.**

# LESSON - 2
# STATISTICAL ENQUIRY - COLLECTION OF DATA

**OBJECTIVES :**

After studying this lesson you should be able to understand the following :

1.  What is statistical enquiry.

2.  How to collect the data.

3.  Statistical System in India.

**STRUCTURE OF THE LESSON :**

2.1  Introduction

2.2  Statistical Inquiry - Methods

2.3  Primary data - Methods of Collection of Data.

2.4  Drafting the Questionnaire for Collection of Data

2.5  Sources of Secondary data

2.6  Differences between Primary and Secondary Data.

2.7  Statistical System in India

2.8  Exercise

## 2.1  INTRODUCTION :

Statistical enquiry means search for knowledge. It is also known as statistical investigation or survey. Statistical investigation is a technical job which requires specialized knowledge and skill. It uses statistical methods. Statistical investigation provides answers to various management prob lems.

'Griffin' defined statistical enquiry as "Statistical enquires have always required considerable skill on the part of the statistician, rooted in a broad knowledge of the subject matter area and combined with considerable ingenuity in over coming practical difficulties.

Statistical enquiry is two types.

```
                    Statistical Enquiry
                            |
        +-------------------+-------------------+
        |                                       |
   Enquiry for                            Enquiry for
   Special Purpose                        General Purpose
```
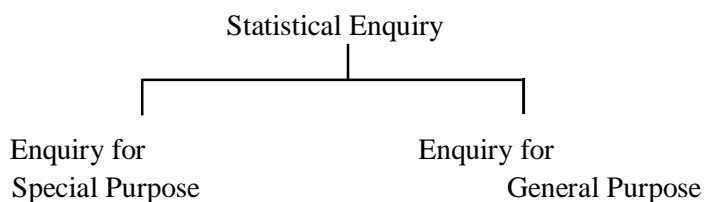
**Enquiry for Special Purpose**: It related to that field in which we have special mission to fulfil.

**Enquiry for General Purpose**: It may relate to the fulfilment of any objective under consideration for which data are collected.

**2.2    STAGES IN STATISTICAL INQUIRY :** A statistical enquiry is a comprehensive process which passes through the following stages :

**2.2.1 Planning the Statistical Inquiry :** A proper planning is essential before a statistical investigation or inquiry is conducted. Careful planning of statistical investigation is essential to get the best results at the minimum cost and time. Following points should be considered in statistical inquiry.

1. Objective of the inquiry should be clear

2. Scope of the inquiry should be determined

3. Scope of the information should be decided

4. Unit of data collection should be defined

5. Source of data collection should be decided

6. Method of data collection should be decided

7. Reasonable standard should be fixed

**2.2.2 Execution of an Inquiry :**

Execution should follow through out the following steps.

i)      Collection of data

ii)     Editing the data

iii)    Presentation of data

iv)    Analysis of data

v)     Interpretation of data

vi)    Presentation of final    report

**i)   Collection of Data :**

The first step in the conduct of an investigation or inquiry is collection of data. The person who conducts the inquiry is known as an investigator. The persons from whom the information is collected are known as respondents. The persons who help the investigator in collecting data are called enumerators. The sources of collection of day may be primary or secondary. The data may be internal or external.

**ii)   Editing the Data :**

Editing the data refers to detect possible errors and irregularities committed during the collection of data. If the data are not edited then it may lead to wrong conclusions. Therefore, editing is essential to arrange the data in order.

**iii)  Presentation of Data :**

The collected data is presented through tables, series graph or diagrams. The classified data is to be presented in such a fashion that it becomes easily intelligible or understandable.

#### iv) Analysis of Data :

Once the data is collected and presented, the next step is that of analysis. The main objective of analysis is to prepare data in such a fashion so as to arrive at certain definite conclusions.

#### v) Interpretation of Data :

The next stage in statistical investigation is interpretation of data. It means to draw out conclusions from the collected and analysed data.

#### vi) Presentation of Final Report :

The final report is prepared with the analysed data.

### 2.3   METHODS OF COLLECTING PRIMARY DATA :

Primary data is one which is collected by the investigator for the first time. it is also known as first hand information. For instance if the extent of malaria in the city is to be computed, then the information regarding the facts collected by the investigators would be termed as primary data. In India agencies like National Sample Survey (NSS), State Level Economic and Statistical Departments collect Primary data. Following methods may be used to collect the primary data.

Methods of Collecting Primary
Data

| Direct Personal Interview | Indirect Personal Investigation | Information through Correspondents | Questionnaire | Schedules | Telephone |
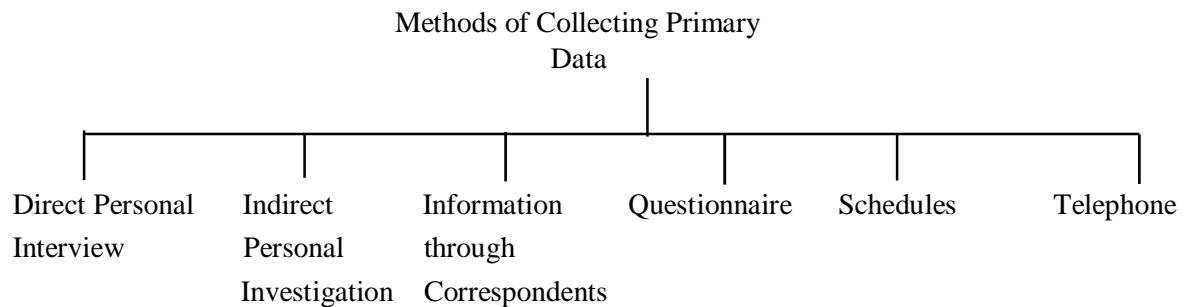
### 2.3.1 Direct Personal Interviews :

Under this method of collecting data, there is a face-to-face contact with the persons from whom the information is to be obtained. The interviewer asks them questions pertaining to the survey and collects the desired information.

**A. Merits :** The advantages of personal interviews are

    i)      Response is encouraging because of personal approach
    ii)    The information obtained by this method is likely to be more accurate.
    iii)   It facilitate to collect supplementary information about the informant's personal characteristics.
    iv)   This system avoids inconvenience and misinterpretation on the part of the informants.

**B. Demerits :** Important limitations of the Personal Interview method are :

    i)      It is very costly method of collection of data, if the number of persons to be interviewed is large and they are spread over a wide area.
    ii)    The chances of personal prejudice and bias are greater under this method.
    iii)   More time is required for collecting information by this method.

### 2.3.2 Indirect Oral Investigation :

Under this method of collecting data, the investigator contacts third parties or witnesses capable of supplying the necessary information. This method is generally adopted by Government Committees. This methods is useful when the direct sources do not exist and cannot be relied upon.

### Merits :

Following are the important merits of indirect oral investigation.

i.   The investigator can take the help of expert enumerators to collect the data.

ii.  Intensive and extensive investigation is possible

iii. It is economical

**Demerits :** If the enumerator is not skilled then wrong data may be collected. The chances of personal bias are greater.

### 2.3.3 Information through Correspondents :

Under this method the investigator does not collect the information from the persons concerned directly. He appoints local agents in different parts of the area under investigation. These local agents are called correspondents. These correspondents collect the information and pass it on to the investigator from time to time.

### Merits :

i.   It is cheap and economical

ii.  It covers large area iii. It is useful when regular information is required.

### Demerits :

i.   The chances for personal base are greater.

ii.  The collected data may not be uniform.

### 2.3.4 Questionnaire Method :

In this method, the necessary information is collected from the respondents through a questionnaire. A questionnaire is a set of questions relating to the enquiry.

### Merits :

i)   Wide coverages is possible

ii)  It is economical because no enumerators are required.

iii) It saves time.

iv)  It is unaffected by the personal bias

### Demerits :

i)   It is costly because enumerators have to be paid

ii)  It is time consuming

iii) It can be employed only by big organizations.

## 2.3.5 Schedules :

Another method of collecting information is that of sending schedules through the enumerators or interviewers. The enumerators contact the informants, get replies to the questions contained in a schedule and fill them in their own handwriting in the questionnaire form. This method is free from most of the limitations of the mailed questionnaire method.

### Merits :

i) It can be adopted in those cases where informants are illiterate.

ii)There is very little non-response iii. Information received is more reliable as the accuracy of statements can be checked by supplementary questions wherever necessary.

### Demerits :

    i)     It is costly
    ii)    It is time consuming
    iii)   It requires trained enumerators
    iv)    It can be employed only by big organizations.

## 2.3.6 Telephone Interview :

The investigator may also obtain information on telephone. For instance the television viewers may be asked to comment on certain programmes on phone.

### Merit :

    i)     This method is less expensive.

    ii)    The scope is wide.

### Demerits :

    i)     A limited group can be approached
    ii)    Very few questions can be asked
    iii)   The respondents may give vougue and reckless answers.

## 2.4   DRAFTING THE QUESTIONNAIRE :

Before framing the questionnaire it is essential to frame in detail the data which we desire from the answers to questionnaire. The success of the questionnaire method of collecting information depends largely on the proper drafting of the questionnaire. The following general principles may be helpful in framing a questionnaire.

    a.    The questions should not be too lengthy

    b.    A decent paper and printing is to be chosen.

    c.    The questions asked should be well worded and should not be ambiguous.

    d.    The questions asked should be in proper sequence.

    e.    Irrelevant questions to the study should be avoided.

    f.    Questions should be free from personal bias and they should not injure the writing wok.

g.    Necessary instructions and definitions should be given.

h.    Questions involving the mathematical calculations should be avoided.

i.    There should be guarantee to keep the answers secret and to use them only for the purpose of said investigation.

j.    The covering letter.

## 2.4.1 Model Questionnaire :

**STUDY OF CHANGING PATTERN OF CORPORATE MANAGEMENT IN INDIA**

1.    Name of the Company         ...................................
2.    Registered Address ...................................
3.    Line of Business ...................................
4.    Total Paid up Capital ...................................
      (a) Number of Shares ...................................
      (b) Class of Shares    ...................................
5.    Shares held by Government financial institutions including Banks.
      Ans...................................
6.    System of Management adopted by your company.
      Ans...................................
7.    Where you a managing agency company or a company managed by a managing agent, please describe the activity in which the erstwhile managing agency company is now engaged in viz.
      (a) Trading    (b) Manufacturing         (b) Processing
      (d) Wound up or not            (e) Investment                    (f) Consultancy Service
      (g) Miscellaneous

8.    If the managing agency company is currently engaged in consultancy services - please state.
      (a) Whether it is rendering service only to the erstwhile managed company ?
      Ans...................................
      (b) Whether its consultancy service can be availed of by other companies?
      Ans...................................

        Please elaborate the services rendered by the consultancy company and number of qualified expects on pay rolls.

      Ans. ...................................
      (c)    Would you suggest any regulation of the consultancy services companies? If so how?
      Ans...................................

9.    (a) Do you believe that after the abolition of the managing agency system, a vacuum created in the management pattern and companies are finding it difficult to have suitable managerial to manage the companies ?
      Ans...................................

(b) If the answer is 'Yes' what in your opinion should be done to develop the managerial talents in the company ?

Ans....................................

10.      (a) Do you think that the provisions of the Companies Act, in relation to management of companies are very cumbersome and that management has to devote more time to com- ply with different legal requirements than to actual management of the company ?

Ans....................................

(b) If answer is 'Yes' what is your opinion are the cumbersome provisions ?

Ans....................................

(c) Do you think that these provisions are dropped or made less strict there would be no mismanagement by those in charge?

Ans....................................

11.      (a) Is there any labour participation in the management of your company ?

Ans....................................

(b) Is it possible in India for labour to participate in management ?

Ans....................................

(c) If the answer is 'Yes' what suggestion you would make for such participation? i) Labour representatives ii) Others.

(d) Do you think that to give them some representation on the Board, employees will have some share holding in the company.

Ans....................................

12.      Make your comments on the law relating to management of corporations in general and suggestions to make ti more efficient or effective.

13.      Please supply one copy of :

(a) Articles of Association

(b) Memorandum of Association (c) Latest Annual Report.

### 2.4.2    Differences between Questionnaire and Schedule :

| Questionnaire | Schedule |
|---|---|
| This method of collecting data can be easily adopted where the field of investigation is very vast. | This method can be adopted where the field of investigation is not very vast |
| It is less expensive | It is more expansive since it required trained staff |
| This method is useful only when informants are literate people. | This method is useful even the informants are illiterate people |
| It involves some uncertainty about the response | There may be no such uncertainty because of direct contact with informants |

## 2.5 SOURCES OF SECONDARY DATA :

The data which is not first hand information (primary data) is known as secondary data. Sometimes it is not possible to collect first hand information for want of resources in terms of money, time, etc., in that situation secondary data is used. This data is mainly classified into two categories.

These are

a) Published data

b) Unpublished data.

### 2.5.1 Published Data :

The published data may be obtained from various International, National and Local Publications. Following are the main sources of Published Data.

### i) Internal Publications :

Certain International Institutions publish reports from time to time regard ing economic matters which are of great significance e.g. Annual Report, Balance of Payments published by IMF, Annual Reports of International Labour Organization (I.L.O.) or by World Bank (I.B.R.D.) etc.

### ii) Official Publications of Central and State Governments :

Generally State and Central Governments collect information regarding important economic variables like national income, savings, investment, employment, etc., and publish it after regular intervals e.g. Report on Currency and Finance, RBI Bulletin published by RBI, Census report published by Census department, Statistical Abstracts are published by every state government at State level. The data published by Planning commission is also called Secondary data.

### iii) Committee Reports :

Sometimes the government appoints survey and enquiry commissions to get the expert views on matters of great importance e.g. Reports of Public Accounts Committee of Lok Sabha.

### iv) Newspapers and Magazines :

The newspapers like the Financial Express, The Economic Times and certain Periodicals like Economic and Political Weekly, Capital, Commerce, Money, etc. Publish the data regarding economic variables.

### v) Individual Research Scholars :

The various reports of research scholars and research institutions also contain data of economic significance.

### 2.5.2 Unpublished Data

When the data are collected by someone but which are          not published and are taken by other persons for his investigation, they are known as Unpublished Secondary Data e.g. reports of trade unions, cooperative societies, reports prepared by private investigation companies etc.
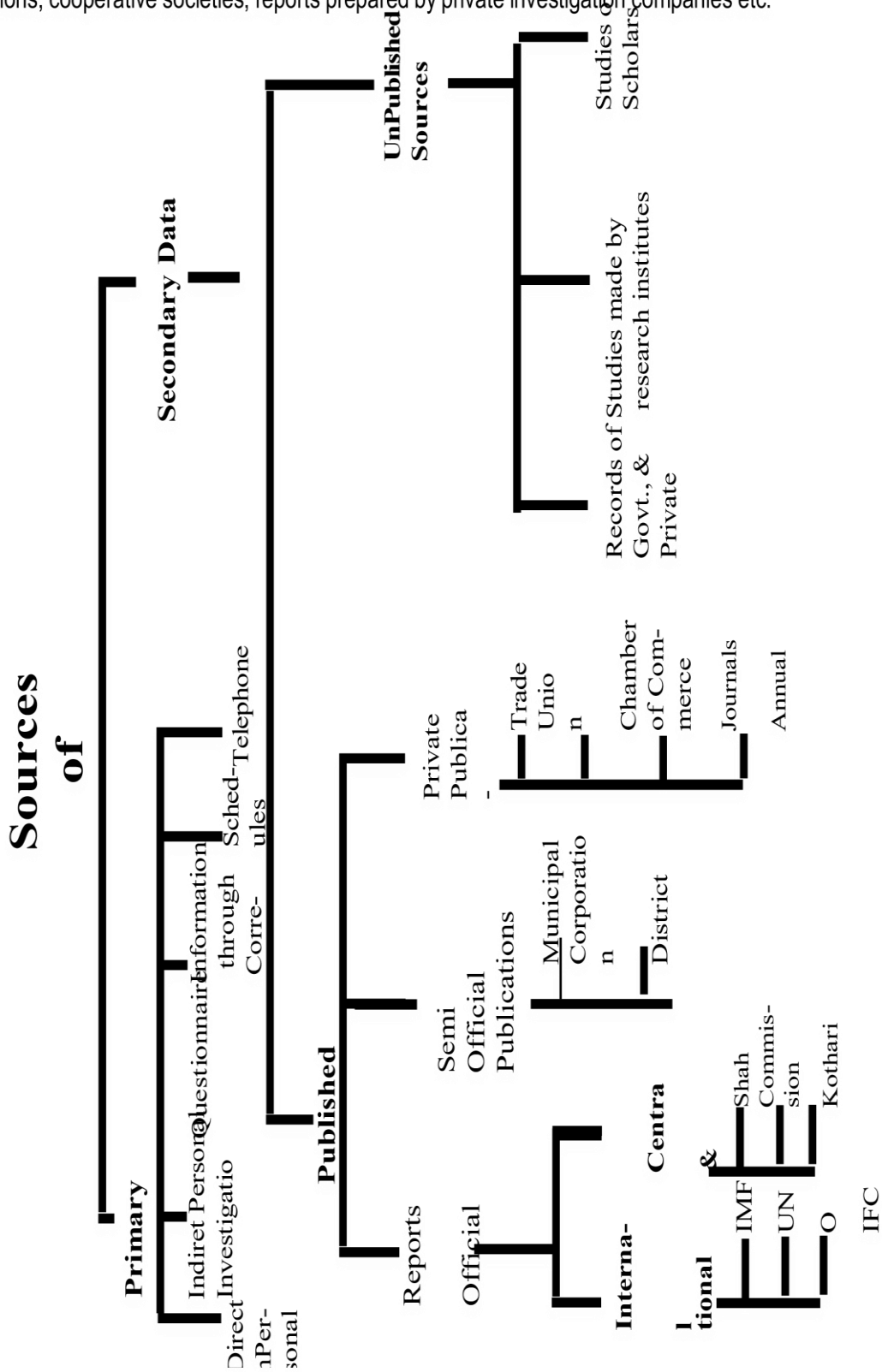
## 2.6 DIFFERENCES BETWEEN PRIMARY DATA AND SECONDARY DATA :

The investigator must decide whether he will use primary data or secondary data in his investigation. While choosing between the two types of data, following considerations should be kept in mind:

i)      Nature and scope of the inquiry

ii)     Availability of financial resources

| Basis | Primary Data | Secondary Data |
|---|---|---|

**1. Cost Factor  :** Needs more funds  Needs comparatively less funds

**2. Source** Investigating Agency collects the data Some other investigating agency    collects    it for its own use.

**3. Time Factor** Requires longer time for collection   Requires less time for collection.

**4. Reliability** More reliable and suitable to the Less reliable and suitable as some- one **and** enquiry because the investigator else has done that job of collection

**Suitability**   himself collects it.        which may not serve the purpose

**5. Organization** Requires elaborate organization No No need of any organizational set up. **n Factor**

**6. Precautions**  extra precautions are required Secondary data need more care and
                  Attention

Availability of time Degree of accuracy desired.

the status of the investigator i.e. Individual or corporation or government etc.

In actual practice most of the statistical analysis rests upon the secondary data. Primary data is used in those cases only where the secondary data does not provide an adequate basis for the analysis.

## 2.7   STATISTICAL SYSTEM IN INDIA :

A national statistical system is required to organize the collection, compilation and publica tion of statistics as important aspects of national life regularly. The system determines the nature, scope and coverage of the statistics to be collected. The national statistical system coordinates the work of the various statistical offices in the country.

### 2.7.1  Types of Statistical System :

A Statistical system can be evaluated from various angles but according to the degree of centralization there are five types of statistical systems which are given below :

1. Totally decentralized system

2. Minimum Coordination system

3. System decentralized by subject with co-ordinating Agency.

4. System with a central office for general statistics and a co-ordinating agency.

5. Centralized system.

## 2.7.2 Indian Statistical System :

Systematic data collection in India started only with the advent of British rule. Before 1947, no serious attempt was made in our country to collect regular and reliable statistics. The present system of statistical organization is decentralized in nature. At present each ministry in the centre has at least one statistical unit.

Thus the present statistical system in India is decentralized one, where the authority and responsibility for collection of statistics is divided between the Central Government and the State Governments on a subject-wise basis. The Central government acts as the co-ordinating agency for presentation of data on an All-India basis. At the Centre, the Central Statistical Organization (CSO), a technical wing of the Department of Statistics located in the Cabinet Secretariat now shifted to the Ministry of Planning, New Delhi, acts as a co-ordinator at the national level of all the activities of the Central and State statistical agencies. At the state level, the State Statistical Bureaus attached to various departments in various State Governments, are charged with responsi bility of co-ordination of all statistics at State Level.

### 2.7.3    Statistical Organization at the Centre :

The Ministry of Statistics and programme implementation is the apex body in the official statistical system of the country. The ministry includes the following.

### A) Central Statistical organization (CSO)  :

The CSO is located in New Delhi. it is responsible for formulation and maintenance of statistical standards. Its functions are as follows :

i) Perform work relating to National Accounts, Industrial Statistics, Consumer Price indices etc.

ii) Conduct of economic census and surveys.

iii) Training in official statistics iv. Coordination of statistical activities under taken within the country and liaising with international agencies in statistical mattes.

The CSO supplies statistical data in the following publications.

1)    UN Statistical Year Book

2)    Un National Accounts Year Book

3)    UN Demographic Year Book

4)    UN Monthly Bulletin of Statistics

5)    Statistical Year Book of ECAFE

6)    Statistical News Letter, etc.

### B. National Sample Survey Organization (NSS) :

National Sample Survey (NSS) was set up in 1950 for conducting large-scale surveys to provide data for estimation of national income and related aggregates especially in the unorganized sector of the economy and for planning and policy formulation. It carries out annually socio- economic surveys covering various aspects of population. Now its personnel strength is abut 6000 in over 170 offices spread throughout the country.

The NSSO is headed by the Chief Executive officer who is also Member Secretary of the governing Council. Its head quarters are in Calcutta and Faridabad. Its activities are as follows.

i)   Survey design

ii)  Field Operations

iii) Processing of data collected and reporting of the results.

iv) The role of NSSO in agricultural statistics is to provide technical guidance to states for conducting crop estimation surveys and to keep continuous watch on quality of crop statistics collected by the state Governments.

v)  The NSSO collects on monthly basis retail price data from selected shops and markets.

vi) Price indices for urban non-manual employees based on these data are compiled and published.

Survey results are published in the form of reports. About 480 reports are available in printed form. NSSO started a quarterly Journal 'Sarve Kshana' from July, 1977. It presents most of the results of NSSO.

At present each Central Ministry has statistical units which are responsible for collection, and compilation of statistics relating to its subject. Important statistical units of the main Central Ministries are as follows.

**Ministry of Planning :** There are four apex bodies, statistical units responsible for co-ordinating and administrative functions related to the collection of statistics in the country by different departments. These units are -

i)   The Central Statistical Organization (CSO).

ii)  National Sample Survey Organization (NSSO)

iii) Computer Centre.

iv) Programme Evaluation Organization.

The CSO and NSSO because of their vital importance have already been discussed.

**Ministry of Home Affairs :** Thirty statistical units are attached to this Ministry. The main publications of this office are :

i)   The Census of India Reports
ii)  Vital Statistics of India (Annual) and
iii) Indian Population Bulletin (Biennial).

**Ministry of Agriculture and Co-operation :** 44 statistical units are attached this ministry. The most important unit is :

Directorate of Economics and Statistics: To compile and publish agricultural statistics on All-India

basis this Directorate was established in 1947. The data covered relate to agriculture, live stock, fishery and forestry. The data are collected monthly by the State Governments. The directorate also serves the Central Government in an advisory capacity.

The important publications of the Directorate are

**Annual Publications :**

1.   Indian Agricultural Statistics

2.   Estimates of Area and Production of Principal crops in India

3.    Indian Agriculture in Brief 4. Indian Livestock Statistics

5.    Indian Forest Statistics.

6.    Agricultural Prices in India

7.    Agricultural Wages in India.

8.    Bulletin on Food Statistics

9.    Tea Statistics.

10.   Coffee Statistics.

## Monthly Publications :

Agricultural Situation in India.

## Weekly Publications :

1. India Livestock Census

2. Indian Crop calendar

3. Bulletin on Commercial Crops

## Ministry of Commerce :

Eight statistical units are attached to this ministry important among them are

## Directorate General of Commercial Intelligence and Statistics :

It was set up in Calcutta in 1895 and the central statistical office was responsible for the collection, compilation and publication of important all-India statistical series till the Second World War. With the formation of statistical units in the various Ministries many of the former functions of this office were transferred to the appropriate Ministries. It's now responsible for commercial intelligence and foreign trade statistics. It's main publications are :

1. Indian Trade Journal (Weekly)

2. Indian Customs and Central Excise Traffic Vols I and II(Annual)

3. Annual Statement of Foreign Sea borne Trade of India.

4. Statistics of Maritime Navigation of India (Annual)

5. Accounts relating to the Inland (Rail and River borne)

6. Trade of India (Monthly)

7. Monthly Statistics of Foreign Trade of India by Country and Currency Areas (Vol I and II Monthly)

**Office of the Chiel Controller of Imports and Exports :** This office publishes annual statistics on imports and exports (Annual Bulletin of Statistics of imports and exports) and annual reports on (Annual Administrative Reports) and weekly reports on licenses relating to industrial exports an imports.

**Ministry of Labour :**

The Labour Bureau was established in 1946 in the Ministry of Labour and Rehabilitation. It collects, compiles and publishes statistics of employment in respect of factories, mines, plantations, shops, commercial establishments etc., on an all-India basis. For the formula tion of labour policy, it provides data after conducting research into the specific problems of labour. It brings out pamphlets on various aspects of labour legislation. it is also responsible for the construction and publication of consumer price index numbers, for industrial agricultural and rural labour.

Its regular publications are :

1. Indian Labour year Book (Annual)

2. Large Industrial Establishments (Annual)

3. Statistics of Factories (Annual)

4. Report on Working of the Minimum Wages Act (Annual)

5. Working of the Trade Unions Act (Annual)

6. Indian Labour Journal (Monthly)

**Ministry of Industrial Development :**

The ministry has seven statistical units. Main statistical unit is the office of the Economic Adviser to the Government of India which was established in 1938. Prior to the setting up of the CSO, it is functioned as the central co-ordinating authority in the field of statistics for the Government of India. Now it maintains wholesale price indices and price data in general and acts as the co-ordinator between various statistical units of the ministry. Its regular publication is Monthly Statistics of Production of Selected Industries.

Besides this, the Development Commission Small Scale Industries publishes yearly, monthly and half yearly reports on the development of small scale industries.

**Ministry of Defence :**

The Army Statistical Organization (ASO) was set up in 1947 under the Ministry of Defence. it performs the following functions:

    i)      Maintenance of basic statistical records and the regular computation and supply of data regarding personnel, vehicles, armament, equipment, animals and accommodation etc.,

    ii)      Control of reports and returns coming from Army and Command Headquarters.

    iii)      Technical advice on statistics in the army.

    iv)      Design, conduct and analysis of sample surveys, experiments and investigation.

The ASO has one of the largest installations in India for mechanical tabulation of data. A research unit is concerned with the development of survey methods and operations research techniques.

**2.7.4 Statistical Organization in the States :**

The apex statistical agency in each State or Union territory is a Statistical Bureau known by different names such as Directorate of Economics and Statistics, Bureau of Economics and Statistics, Directorate of Statistics and Evaluation, Economic and Statistical Organization, Economic and

Statistical Advisory to the State government etc. These are generally under the administrative control of the Finance or Planning Department of the concerned state. The main functions of the State Statistical Bureau are :

1. Systematic Collection, Compilation, analysis, co-ordination, and interpretation of the statistics relating to the States.

2. To act as an advisory body on economic issues referred to it.

3. Organizing and conducting special enquiries and field surveys.

4. Liaison between statistical organization of the Centre and other States.

5. Publication of an annual Statistical Abstract and monthly, quarterly bulleting including all essential statistics of the State.

6. Compilation of economic indicators and State Income Estimates.

7. Statistical Work relating to planning.

8. Publication of Socio-Economic Surveys of the State to be presented in the Budget Session of the State.

### 2.7.5 Non-Governmental Statistical Organization :

The following non-government organizations are working in the country.

1. Indian Statistical Institute, Calcutta.

2. Institute of Agricultural Research Statistics, New Delhi

3. Statistical Department of the Reserve Bank of India.

4. National Council of Applied Economic Research, New Delhi.

5. Institute of Economic Growth, Delhi.

6. Institute of Foreign Trade, New Delhi.

7. Gokhale Institute of Economics and Politics, Pune.

8. Tata Institute of Social Sciences, Bombay.

9. Institute of Labour Research, Bombay.

10. Economic Department of the Reserve Bank of India.

11. Universities in India.

### 2.7.6 National Statistical Commission (N.S.C.) :

The commission after examining the present system of collection of dissemination of statistics relating to different sectors of the economy adopted a five fold approach to bring about improvements.

1. Reform in the administrative structure of Indian Statistical System and upgrading its infrastructure as to ensure its autonomy.

2. Improvement of present system of collection of data.

3. Exploration of alternative techniques, in relation to the existing statistics, if the present system for collecting data is under strain for whatever reasons.

4. Identification of new data series that may be generated in keeping pace with the expanding economy.

5. Evolution of appropriate methodologies for collection of data in relation to new data requirements.

## 2.7.7 Features of National Statistical Commission :

1. NSC has produced a comprehensive report on all aspects of the Indian Statistical System.

2. The commission's approach for improving and strengthening the statistical base has taken the obvious form of recommending about 10 census studies, over 60 types of sample studies and series of other data gathering activities many of which would be fresh efforts.

3. It would cover not only myriad segments of unorganized or informal sectors but also organized sectors like private corporate sector, NBFC's and even registered factories sector.

4. The Commission has advised the government to exercise caution on enthusiasm shown by government departments to engage private sector organizations as data collection agencies.

5. It has addressed all issues in their entirety.

## Defects of National Statistical Commission :

1. There is no sign of any innovation in it.

2. There is no vision of the possible course of changes taking place in Indian polity and the economic structure.

3. If NSC report not focuses on requirements with developmental objectives.

4. Commission has failed to give proper attention to inadequacies in the estimation of domestic saving and investment.

## Suggestions :

1. The commission should emphasize on building of Regional Accounts not only at states level but also at an invariant regional grouping states.

2. There is need to break new ground in Industrial Statistics

3. NSC could have suggest the establishment of a system to monitor the progress made in new industrial investment taking place in private sector.

4. There is need to track progress in foreign Direct Investments through the requirement of regular data on projects implemented under FDI.

5. There is need for evolving data on lead in economic indicators.

**Appraisal of Indian Statistical System :**

The National Statistical System covers a wide spectrum of national accounts statistics, industrial

statistics, export and import statistics, labour statistics, vital statistics, agricultural statistics, environmental statistics, meteorological statistics etc. The statistics are collected under collection of statistics Act, 1953 mainly for the Annual Survey of Industries conducted by the National Sample Survey Organization (NSSO) and Census Act, 1948.

All is not well with the National Statistical System (NSS), despite the recommendations of the Review Committee in 1979 for making improvements in the system. The Government of India had set up a National Advisory Board on Statistics on 1982, however, the statistical system is suffering from various deficiencies and gaps.

In order to revamp the statistical system in the country, the Government of India has taken two policy decisions:

1. The government of India has borrowed Rs. 850 cores from the World Bank for Revamping the national statistical system in order to bring it at part with the international standards.

2. The Government under the Ministry of Agriculture has set up a National Crop Forecasting Centre (NCFC) for preparing crop forecasts on scientific lines and enable the Government to take strategic decisions on the price front.

**Summary :**

The Indian Statistical System is still in the process of evolution. With a view to providing a sound statistical base and developing a system of continuous flow of information, the Planning Commission has continuous flow of information. The Planning Commission has constituted two committees, namely, (i) Standing Committee for Improvement of Data base for Planning and Policy Making and ii) Standing Committee for Improvement of Data Base for Decentralized sectors, consisting of members from government and non-government organizations.

## 2.8  EXERCISE :

### A. Short Answer Questions :

1.   How to plan statistical inquiry

2.   What is primary data.

3.   Explain direct personal investigation

4.   What is meant by C.S.O.

5.   What are the functions of NSSO.

### B. Essay Questions :

1.   What are the sources of Collection of data.

2.   Explain differences between Questionnaire and Schedule.

3.   What are the differences between primary data and secondary data.

4.   Explain the Statistical System in India.

**Dr. K. Kanaka Durga.**

**OBJECTIVES :**

After going through the lesson you will be able to understand the following:

1. Presentation of data and its methods.
2. Classification of data, its need, types and methods.
3. Types of statistical series.
4. Tabulation of data, types of tables and rules for tabulation.

**STRUCTURE OF THE LESSON :**

## 3.1    INTRODUCTION :

Classification and tabulation of data occupy an important place in Statistics. Unless data are classified properly and tabulated attractively and meaningfully, they won't serve the purpose. In this lesson, all aspects relating to classification and tabulation are discussed. Further, importance of data presentation and statistical series are also discussed.

## 3.2 PRESENTATION OF DATA :

After the data have been collected and examined, they will have to be presented in a systematic manner either in their raw form as they emerge after editing, or they will have to be statistically treated before their final presentation to the people at large. Generally, data of a simple nature are presented in the form in which they emerge after collection and editing. But data of a more complex nature have to be treated statistically prior to their interpretation. The manner of presentation is very important. If data is not properly presented, it fails to attract due notice and its chief features are not adequately noticed. The data should be so presented that it may be within easy grasp and be swiftly available for easy reference. This is done effectively by graphical or pictorial methods. Whatever method may be employed the chief aim should be to enable one to grasp easily and readily significant proportions, differences or trends in the data.

**3.2.1 Methods of presentation :** The following methods of presentation are commonly used in Statistics:

### 1. Presentation in the form of statements :

Presentation of data in the form of a statement consisting of text and figures is not always effective. It requires careful reading of the text before one is able to understand it. Then it has to be read over again and again as many times as one requires particular information. The main object of Statistics is to simplify complexities. On this score this method does not come up to the mark. The following abstract, taken from the 2007 Wipro Company's report of quarter 2, is an example of presentation of data in the form of statement.

"Wipro has reported a 35 per cent year-on-year revenues for the second quarter ended September 30, 2007 to Rs.4, 785 crore. The net profit stood at Rs.824 crore against Rs.700 crore in the corresponding quarter last year, an increase of 18 per cent. The company has announced an interim dividend of Rs.2 per share. Wipro's Global IT services and products revenue grew only 9.7 per cent sequentially to $796.5 million."

### 2. Presentation in the form of classified statements :

When data are of such a nature that they can be broken into two or more parts according to their distinguishing features they may be so presented. A part of information contained in the clearance of Special Economic Zones issued by the government of India is as follows:

"The center cleared 7 new Special Economic Zones (SEZs),

they are—

One SEZ (ITeS) (TCS)
West BengalTwo SEZs
Tamilnadu

One SEZ (ITeS), Indore

Mandhya Pradesh

One Malwa IT Park Ltd, Bangalore

Karnataka
One Perfect IT SEZ Pvt. Ltd, Noida Uttar Pradesh One Calica Construction Impex Pvt, Ltd, Gandhinagar - Gujarat".

The advantage of this mode of presentation is that figures which are considered as significant may be made to stand out prominently away from the statement. Explanatory notes may be included in the text of the statement. Isolating the figures from the statement makes the data more readily assailable,

and avoids chances of confusion.

## 1. Presentation in the form of tables :

This method of presentation makes the data more swiftly understandable as a mass of complex data is broken into several classes and consigned to appropriate columns in the tables. The title of the table gives brief account of its contents, and if the title is carefully selected it may become sufficiently self-explanatory. It gives the entire information intended to be conveyed in a brief and precise manner which it is very easy to scan. Particular attention can be invited to certain facts and figures by stating the facts in footnotes, and the figures in bold letters.

## 2. Diagrammatic and Graphic Presentation :

This method is generally used as a visual aid, and is gradually coming into prominence. Its importance is being recognized as an effective mode of presenting data.

## 3.3 CLASSIFICATION OF DATA :

Classification is the process of dividing things into different classes or sequences according to the affinities of their character which exist among a diversity of features in them. The process of classification, if carried to its logical conclusion, means that there should be as many classes as items to be classified, because, while they will have some features in common, in several other respects they will be different from each other. Such a classification, then, would lose the very purpose for which it is made. It is, therefore, enough if we classify items according to the object in view. The object of inquiry will determine as to how facts should be separated into groups or classes according to characteristics needed to be studied for the purpose of the investigation.

### 3.3.1 Need for Classification :

The most important function of Statistics is to simplify complexities. A large mass of complex data is not capable of signifying anything unless it is presented in a proper manner, duly divided into groups with respect to some characters which are of a variable nature.

The chief object of classification, therefore, is to rid the data of its complex nature and render it easy to understand. Then, since classification is done according to affinity of character, another object of classification is to separate the similar from the dissimilar, and bring out the distinguishing features. Thus, it enables comparisons to be made and conclusions to be drawn without the necessity of considering directly hundreds of individual numbers. Then, since classification is a logical process, it ensures orderly arrangement of items, which is easy to follow and study further. It, thus, serves as mental and visual aid, and renders tabulation easy.

### 3.3.2 Types of Classification :

Classification may be of two types depending upon the nature of data. If the data is of a descriptive nature, possessing several qualifications which it is possible to classify according to some physical or natural characteristics, it can be classified according to attributes, for example, males and females, Indians and non-Indians, etc.

If, however, data are expressed in numerical quantities, they are classified according to class-intervals, for example, classification of persons according to age-groups as falling between ages 5 to 10, 10 to 25, 25 to 50 years, etc.

### 3.3.2.1 Classification according to Attributes :

When data relate to persons or things laying emphasis on their physical or natural characteristics, they can be classified according to their qualities. The process of classification according to qualities or attributes consists in isolating the similar from the dissimilar. Things or persons possessing the qualities common to each other are placed in one class. Classification according to attributes may be of two kinds.

**Simple Classification:** It is that where only one attribute is studies, for example, classification of

persons according to their sex – males and females; according to literacy – literates and illiterates, etc. When one attributes is observed, it results in classification into two classes – one, consisting of those possessing the attribute, another, consisting of those not possessing the attribute. Thus the two classes are strictly exclusive of each other. A simple classification, where items are classified according to one attribute, forming two sub-classes, is also known as classification by dichotomy.

**Manifold Classification:**

Where more than one attributes are observed, classification may lead to the formation of a number of classes and sub-classes, for example, students are classified as graduate and undergraduate students; among each of the broad classes there are again two sub-classes; males and females, or boys and girls; males and females are further sub-classified as Indians and non-Indians. There is no limit to which we can carry on this process of classification or sub-classification. In the above example the attributes observed are graduate and undergraduate students, their sex and their nationality. More attributes may be observed leading to the formation of further sub-classes.

### 3.3.2.2 Classification according to Class - Intervals :

When data are expressed in numerical characters and it is necessary to make them easy to comprehend, it is sub-divided into classes constructed out of limits formed either arbitrarily or on grounds of convenience. Such a classification is known as classification according to class- intervals. Sometimes, attributes not capable of precise description are defined by numerical notions, for example, tall and short is a classification according to attribute. But who is a tall person? If population of a town is to be studied for some statistical object, it will not serve any useful purpose if we classify the population as infants, children, young, middle aged, and old but, in order to make the data precise, we shall have to adopt some such numerical notations as: below 5 years (infants), 5-10 years (children), 10-35 years (young), 35-55 years (middle aged) and above 55 years (old).

It is necessary to study certain terms which are used in connection with classification according to class-intervals. Firstly, the classes (viz. below 5, 5-10, 10-15, 35-55 etc.) are known as class-intervals. The figures 5, 10, 35, 55 etc. are known as limits of the class-intervals. In the class-intervals given above, the first figures in each of them are the lower limits of the class- intervals and the second figures the upper limits of the class-intervals. The difference between the upper limit and the lower limit of a class-interval is known as the magnitude of the class-interval. If this difference is the same throughout the various classes in the class-intervals, the magnitude is known as uniform, example, 0-5, 5-10, 10-15, 15-20 etc. The difference is 5 in each case. But if the difference changes in various class-intervals it is known as un-uniform magnitude, example, 0-5, 5-10, 10-35, 35-55 and 55-80. Here the difference varies from class to class. In the first two cases it is 5 in each case, in the third it is 25, in the fourth 20 and in the last, again 25. The number of items belonging to each of the classes is called the frequency of the class-intervals. If for each of the class-intervals the frequencies given are aggregates of the preceding frequencies, they are known as cumulative frequencies, otherwise they are known as individual class frequencies or simply frequencies. The frequencies may be cumulated either from the top or from the bottom. The class-intervals are put accordingly.

### 3.3.3 Methods of framing class-intervals :

The method according to which the above class-intervals, viz. 0-5, 5-10, 10-15 etc., are framed is known as the '*exclusive method*'. Here the class-intervals overlap. In assigning items to various classes, the main difficulty which arises is as to what class should items falling on the limits be assigned, for example, whether '5' should be included in the first class or in the second class, and similarly whether '10' belongs to the second class or to the third class. In the exclusive method, an item which is identical to the upper limit of a class-interval is *excluded* from that class- interval and is included in the next class-interval. Hence it is called 'exclusive method'. An item, the measurement of which is exactly '5' will belong to the second class and not to the first, and so on. For all practical

purposes, therefore, the class-interval '0-5' means from '0' to less than '5', '5-10' means from '5' to less than '10' and so on.

There is another method of framing the class-intervals, where the above ambiguity about items identical to a limit of the class-interval is sought to be removed. This method is known as 'inclusive method'. The above class-intervals according to the inclusive method will read as: 0-4, 5-9, and 10-14 etc. To remove difficulty of an item which is not a complete number and falls between the upper limit of a class and the lower limit of the next class, the above class may be expressed according to inclusive method also as: 0-9.5, 5-9.5, 10-14.5 etc. or 0-4.9, 5-9.9, 10-14.9 etc.

It should, however, be noted that whether the upper limit of the first class is expressed as 5, or 4, o4 4.5 or as 4.9, it would always stand for 'less than 5' and the magnitudes of the class- interval will be 5.

### 3.3.4 Class-Intervals with Cumulative Frequencies :

Sometimes class frequencies are not given as individual class-frequencies but as cumulative class frequencies. When frequencies are cumulated, the measurement of class- intervals is also cumulated. Frequencies may be cumulated either from the top or from the bottom. The class intervals are not expressed in usual manner with their lower and upper limits, but only with the upper limits preceded by the word 'below', (or 'less than'), or 'above' (or 'more than') as the case may be according to as the frequencies are cumulated from the top or from the bottom.

Before treating such data statistically, it is necessary to convert them into usual class-intervals and individual class frequencies. The following example shows how frequencies cumulated from the top and from the bottom are converted into usual types of data:

1. Class frequencies cumulated from top

2. Class frequencies cumulated from bottom

| Marks | | Number of Students | Marks | Number of Students |
|---|---|---|---|---|
| Below | 5 | 10 | Above 0 | 55 |
| Below | 10 | 22 | Above 5 | 45 |
| Below | 15 | 37 | Above 10 | 33 |
| Below | 20 | 50 | Above 15 | 18 |
| Below | 25 | 55 | Above 20 | 5 |

The above data converted into usual type of class-intervals and individual class frequencies will read as follows:

| Marks | No. of Students |
|---|---|
| 0 – 5 | 10 |
| 5 – 10 | 12 |
| 10 – 15 | 15 |
| 15 – 20 | 13 |
| 20 – 25 | 5 |

**General Considerations :** It is for the statistician to decide about classifications, but some general considerations need to be taken care of:

1. The classification must be exhaustive. It should be possible to include each of the given values in one or the other class.
2. The classes must be mutually; exclusive i.e. they should not overlap. If, however they have to overlap as in the case of exclusive classes, the statistician must observe the rules of classification applicable to such classification.
3. The number of classes should be neither too large nor too small; for either of the practices is

likely to undermine the purpose of classification, and upset the pattern of distribution of the frequencies. It is not possible to lay down the number of classes which may be applicable to all situations.

4. The magnitude of class-intervals should be uniform, if possible, throughout the classification, and the system of 'open' classes should be avoided.

## 3.4    STATISTICAL SERIES :

According to L.R.Connor, "if two variable quantities cab be arranged side by side so that the measurable differences in the one correspond to the measurable differences in the other the result is said to form a statistical series". In other words, any logical or systematic arrangement of items constitutes a series. When things or attributes are counted, measured, or weighed and placed one after the other in some orderly manner, they are said to form a series.

As discussed in the above pages, series or arrangement of data can be done on the basis of time, space, or some conditions. So far as time series and spatial series are concerned, there is no problem in their formulation. Frequencies can be noted down on the basis of time or space, however, when series are formed on the basis of changes in some condition like age, weight, marks, production etc., and the series can be either discrete or continuous. Let us discuss about them in detail.

**Discrete Series :**

When items are arranged in groups showing definite breaks from one point to another, and when they are exactly measurable, they constitute a discrete series. Items are arranged in ascending or descending order and opposite them the number of times each item occurs is mentioned. In a question in which the maximum marks were six, students secured marks as follows:

| Marks | No. of Students | Marks | No. of Students |
|-------|-----------------|-------|-----------------|
| 1     | 5               | 4     | 7               |
| 2     | 8               | 5     | 6               |
| 3     | 10              | 6     | 1               |

After each marks group 1, 2, 3 and so on there are definite breaks and the students seem to secure exact marks as 1, 2, 3, and not as fractions. Such a series is termed as a discrete, or a broken or a discontinuous series.

**Continuous Series :**

When items are arranged in groups or classes because they are not exactly measurable, they form a continuous series. Items which are capable of precise measurement should either be placed in a series of individual observations or in a discrete series. But when it is not possible to measure them in exact terms, or if it is possible to so measure them but the measurements, they are entered into classes or groups of measurements.

## 3.5    TABULATION OF DATA :

Tabulation of data is the last stage in the compilation of data, and forms the basis for its further statistical treatment. It is a systematic presentation of data in columns and rows. The following are the important definitions of tabulation.

"The logical listing of related quantitative data in vertical columns and horizontal rows of numbers with sufficient explanatory and qualifying words, phrases and statements in the form of titles, headings and explanatory notes to make clear the full meaning, context and the origin of the data" Tuttle.

"Tabulation is the process of condensing classified data in the form of a table so that it may be more easily understood, and so that any comparisons involved may be more readily made" D.Gregory and H.Ward.

### 3.5.1 Objectives of Tabulation :

The following are the main objectives of Tabulation.

1. **To simplify complex data :** In the process of tabulation of data, unnecessary details are avoided and data are presented systematically in columns and rows in a concise form. All tabular data are presented in such a manner that they become more meaningful and can be easily understood by a common man.

2. **To facilitate comparison :** Data presented in rows and columns facilitate comparison. Since a table is divided into various parts and for each part separate sub-totals and totals are given relationship between various items of the table can be easily understood.

3. **To economize space :** Economy of space is achieved by tabulation, as all unnecessary details and repetitions are avoided without sacrificing quality and utility of the data.

4. **To depict trend and pattern of data :** Tabulation of data depicts the trend of the information under study and reveals the patterns within the figures which cannot be understood in a descriptive form of presentation.

5. **To help reference :** When data are arranged in tables with titles and table numbers, they can be easily identified and made use of, as source reference for future studies.

6. **To facilitate statistical analysis :** After classification and tabulation, statistical data become fit for analysis and interpretation. Various statistical measures like averages, dispersion, correlation, etc., can be calculated easily from the data which are systematically tabulated.

### 3.5.2 Types Of Tables :

From the standpoint of usage, statistical tables are of two types:

1. General Tables

2. Summary or Special purpose Tables

1. **General Tables :** These tables contain a mass of detailed information including all that is relevant to the subject-matter. Hence such tables are very large, extending over a number of pages. The main purpose of such tables is to present all the information available on a certain problem at one place for easy reference. They usually find their place in the appendix of reports or special studies of problems.

2. **Summary Tables :** These tables are designed to serve some specific purposes. They are smaller in size than general tables and seek to lay emphasis on some aspect of data. They are generally contained in the text. They are called summary tables because they are brief and are also called derivative tables because the general tables serve as the source from which they are derived or made. They aim at analysis and comparison of data and enable conclusions to be drawn.

### 3.5.3 Forms of Tables :

Tables may be simple or complex to form. Let us discuss about them.

1. **Simple Tabulation:** In this type of tabulation, a table contains information pertaining to only one set of related data and seeks to answer one or more groups of an independent investigation. Observe the following table:

| Marks | No. of students | Marks | No. of students |
|---|---|---|---|
| 0 – 5 | 15 | 15 – 20 | 25 |
| 5 – 10 | 17 | 20 – 25 | 20 |

| 10 – 15 | 22 | 25 – 30 | 13 |

Thus, a simple table has two factors placed in relation to each other.

2. **Complex Tabulation :** In this type of tabulation, a table contains information pertaining to a number of coordinate factors. If there are two coordinate factors, the table is called a double table; if the number of coordinate groups is three it is a case of treble tabulation; and if it is a case of more than three coordinate groups the table is known as multiple tabulation.

In the above table, if the students are further classified into groups according to residence, hostellers or day scholars, it will be a case of double tabulation. If the students falling into each of the two groups – hostellers and day scholars – are classified according to sex, it will be a case of treble tabulation. If they are again classified as belonging to different religions, states, nationalities etc., it will constitute an example of manifold tabulation. More than one factor makes the plan of a table slightly complex, and the larger the characteristics distinguished the more complex the table becomes.

### 3.5.4 Rules and Precautions for Tabulation :

There are no hard and fast rules for tabulation. Experience is the best guide and practice is the best teacher to enable good table to be drawn. The main consideration, however, is that a table should amply fulfill the purpose it is designed to and must make the data readily assailable. With this end in view certain rules of procedure are laid down for the guidance of statisticians.

1. The chief consideration should be to make the table as **simple** as possible, free from all avoidable confusion. Then only it may bring out its chief features and the required information quite easily. Clarity should not be sacrificed at any cost for that is the main function of tabulation. If it is necessary to include a mass of relevant information, it may often be found convenient to break it into two or more tables accompanied by a summary table. Every table must be a unit by itself, dealing with different groups or sections of information. Too many details in a table confuse the eye, and make comparisons and detection of errors more and more difficult.

2. Figures to be **compared** should be placed as near to each other as possible, and absolute figures as well as figures expressed in units of comparison, for example, averages, percentages etc., should be shown for easy comparison. Figures to be compared should be placed in vertical columns as far as possible so that they may be compared easily. Totals to be compared may be given in bold type if it is possible.

3. helpful to those who use the data in forming their own estimates about the reliability of the data.

4. The **ruling** should be such that major items are separated by bold or double lines.

5. For all important and principal heads there should be **separate columns**, and minor heads may be placed in one column which may be called 'miscellaneous'. The miscellaneous column must contain only those items which are not of a widely varying nature.

6. If certain data are **not available** for inclusion in the table this fact must be mentioned in the footnote by giving a suitable 'mark' (like N.A for not available) in the appropriate place where such data ought to figure.

7. The columns should be properly **'ranged'** by putting thousands under thousands and hundreds under hundreds. This gives an orderly appearance to the table.

8. The **arrangement of items** in the table should follow some logical order. They may either be arranged in order of their magnitude, or in alphabetical, geographical, and chronological or in any other suitable arrangement.

## 3.6   SUMMARY :

Classification of data serves the purpose of easy understanding. According to the requirement, data can be classified. The data thus classified should be arranged in a systematic manner called series. Later, the data should be tabulated for easy understanding and viewing. Some rules and precautions, if followed, the tables should be attractive and meaningful.

## 3.7   GLOSSARY :

**Classification** – It is the process of dividing data into different classes or sequences according to the features in the data.

**Series** – It is a logical or systematic arrangement of items.

**Tabulation** – It is a scientific process involving the presentation of classified data in an orderly manner.

## 3.8   SELF ASSESSMENT QUESTIONS :

1. What is classification? Describe the various bases of classification.

2. Explain various types and forms of tables.

3. What are the guiding principles in the construction of a table?


**-Dr. R. Jayaprakash Reddy.**

# LESSON - 4
# DIAGRAMS & GRAPHS

**OBJECTIVES :** By the study of this lesson you will be able to understand the importance and tility of diagrams and various types of diagrams. You will also be able to understand the importanceand utility of graphs and various types of graphs.

## STRUCTURE OF THE LESSON :

## 4.1  INTRODUCTION :

Although tabulation is very good technique to present the data, but diagrams are an advanced technique to represent data. As a layman, one cannot understand the tabulated data easily but with only a single glance at the diagram, one gets complete picture of the data presented. Accordingto M.J. Moroney ,-Diagrams register a meaningful impression almost before we think".

## 4.2  IMPORTANCE OR UTILITY OF DIAGRAMS :

1. Diagrams give a very clear picture of data. Even a layman can understand it veryeasily and in a short time.

2. We can make comparison between different samples very easily. We don't have touse any statistical technique further to compare.

3. This technique can be used universally at any place and at any time. This technique isused almost in all the subjects and other various fields.

4. Diagrams have impressive value also. Tabulated data has not much impression ascompared to Diagrams.  A common man is impressed easily by good diagrams.

5. This technique can be used for numerical type of statistical analysis, *e.g.* to locateMean, Mode, Median or other statistical values.

6. It does not save only time and energy but also is economical. Not much money isneeded to prepare even good diagrams.

7. These give us much more information as compared to tabulation. Technique oftabulation has its own limits.

8. This data is easily remembered. Diagrams which we  see leave  their  lasting impression much more than other data techniques.

9. Data can be condensed with diagrams. A simple diagram can present what even cannot be presented by 10000 words.

## 4.3  RULES OR DIRECTIONS FOR MAKING DIAGRAMS :

While preparing the diagrams we must observe some rules to make these diagrams more impressive and useful.

1. It  must be attractive.

2. Its presentation must be proportionate in height and width.

3. It must be Economical in terms of money, energy and time.

4.    It must be Intelligible.

5.    Scale must be presented along with diagram.

6.    Size of figure should be  such that it may occupy considerable portion of paper.

7.    It must be self-explanatory. It must indicate nature, place and source of data presented.

8.    It must be neat and clean.

9.    Diagrams are of several types. The diagram drawn must be suitable to data.

10.   If some points  are to be clarified, foot notes may be given.

11.   Different shades, colors can be used to make diagrams more easily understandable.

12.   Vertical diagram should be preferred to Horizontal diagrams.

13.   If possible, suitable title may be given.

14.   It must be accurate. Accuracy must not be done away with to make it attractive orimpressive.

## 4.4    LIMITATIONS OF DIAGRAMS :

1.    Diagrams depict only approximate results. Those are not so accurate.

2.    Due to above reasons these can't be put for further analysis.

3.    If scales are different, two diagrams can't be compared.

4.    For false base diagrams, a lay man may not make difference.

## 4.5    TYPES OF DIAGRAMS :

Diagrams can be classified into following categories :

1.    One-dimensional  Diagrams.

2.    Two-dimensional Diagrams.

3.    Three Dimensional Diagrams.

4.    Pictograms or Picture Diagrams.

5.    Cartograms or Maps.

## 4.6    ONE DIMENSIONAL DIAGRAMS :

In this case only the length dimension is given the importance. These diagrams are either Bar or   Line Diagrams.

### 4.6.1   Line Diagrams :

In these diagrams only line is drawn to represent one variable. These lines may be vertical or horizontal. The lines are drawn such that their length is in proportion to value of the terms or items so that comparison may be done easily.

**Example 1.** No. of accidents in a city in a year is given below :

Month : 1 2 3 4 5 6 7 8 9 10 11 12

No of Accidents : 8 12 20 16 10 16 20 14 10 19 16 10

**Solution :** Prepare line diagram.

Scale Y - axis 1 cm. = 2 AccidentsOn graph 1 Div. = 2 Accidents
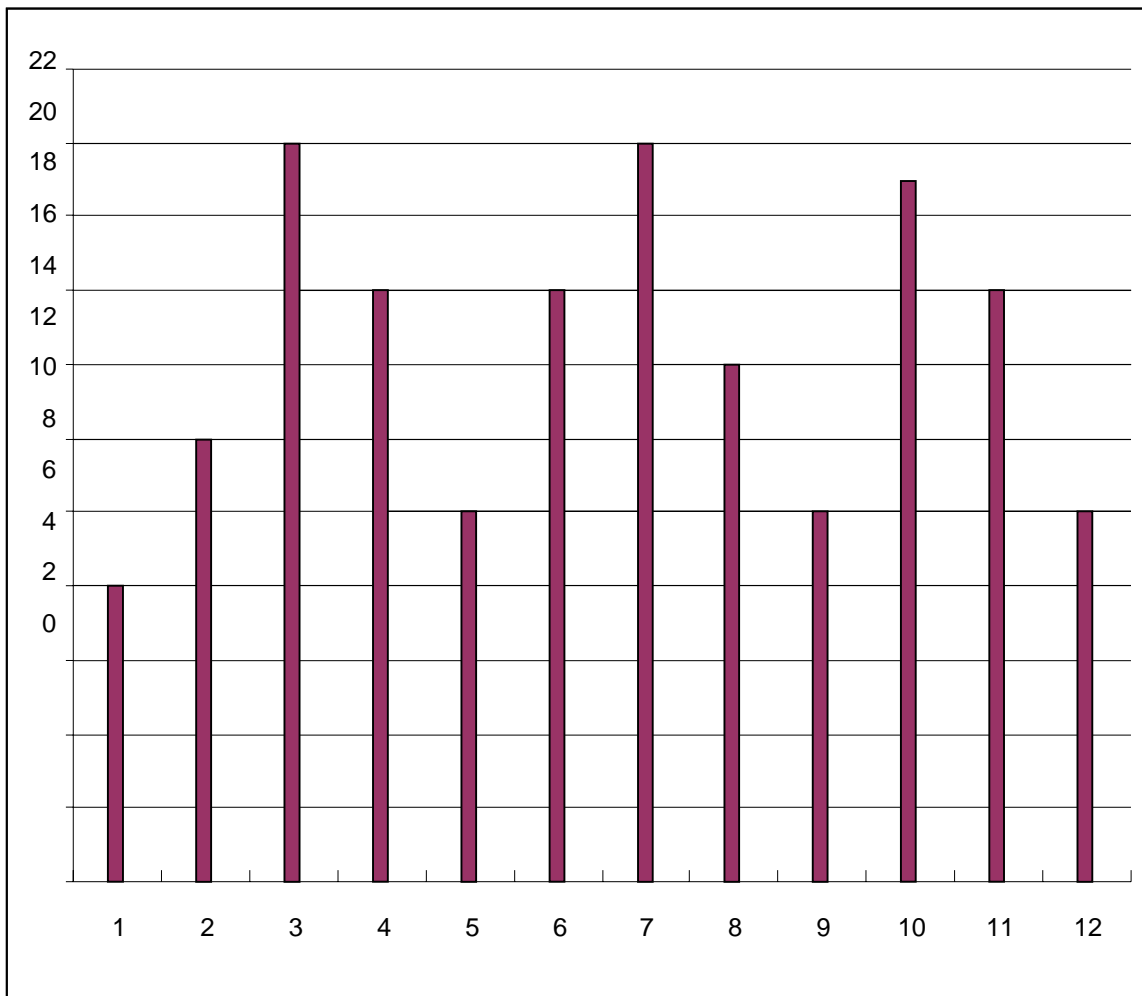


### 4.6.2 Simple Bar Diagram :

Like line diagrams these figures are also used where only single dimension *i.e.* length can present the data. Procedure is almost the same, only the thickness of lines is measured. These can also be drawn either vertically or horizontally. Breadth of these lines or bars should be equal. Similarly distance between these bars should be equal. The breadth and distance between them should be taken according to space available on the paper.

**Example 2.** Average wages of some firms are given below. Represent this by simple Bar Diagram.

Firm : A B C D E F

Average wage : 345 598 540 305 190 150

**Solution :**

Scale Y - axis 1 cm. = Rs. 100On graph 1 Div. = Rs.10



**S.69**
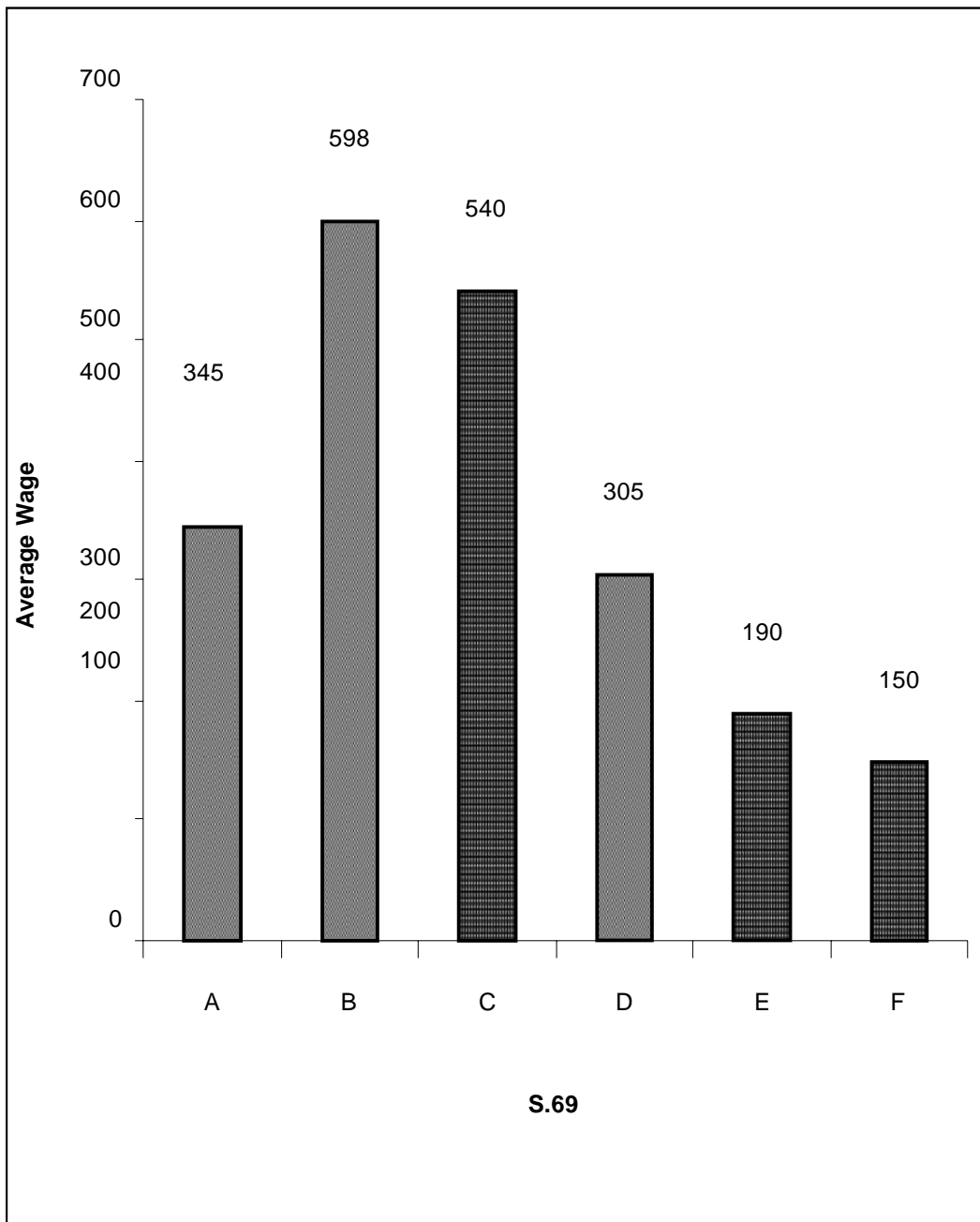
**Example 3.** Present the following data by horizontal bar diagram.

| City | :Jalandhar | Amritsar | Ludhiana | Patiala | Ropar |
|------|-----------|----------|----------|---------|-------|
| Population (Lakhs) : | 8 | 9 | 11 | 7 | 8 |

**Solution :**

Scale x axis 1 cm. = 2 lakhs,on graph 1 Div. = 2. Lakh.
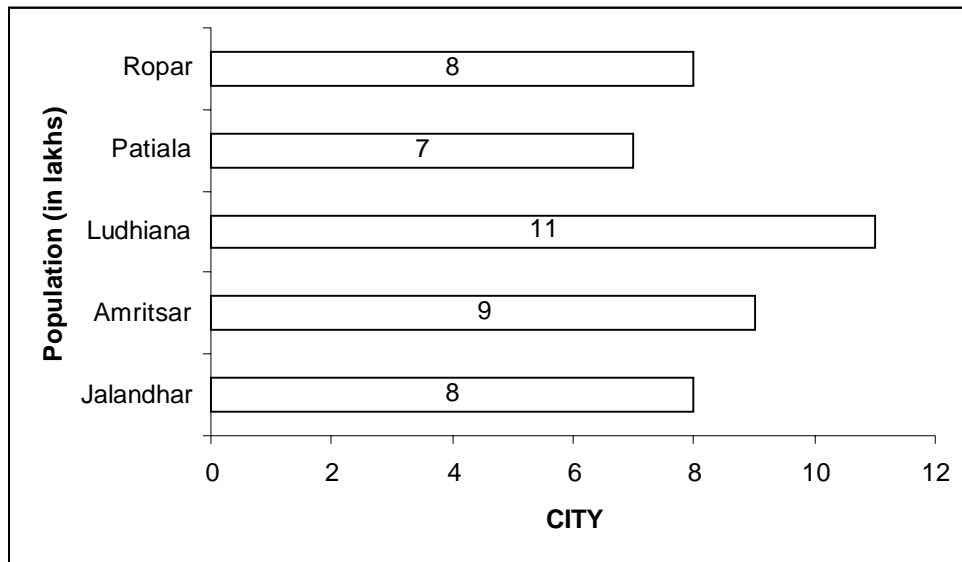


**Multiple Bar Diagrams :**

This diagram is used, when we have to make comparison between more than two variables. The number of variables may be 2, 3 or 4 or more. In case of 2 variables, pair of bars is drawn. Similarly, in case of 3 variables, we draw triple bars. The bars are drawn on the same proportionate basis as in case of simple bars.

**Example 4.** No. of students in Postgraduate classes in a university is given below :

|  | **Science** | **Humanities** | **Commerce** |
|---|---|---|---|
| 2002-03 | 240 | 560 | 220 |
| 2003-04 | 280 | 610 | 280 |
| 2004-05 | 340 | 570 | 370 |

**Solution :**

Here biggest item or term is 610, we should take last term in graph as 650 or 700.

Scale x axis 1 cm. = 100 studentson graph 1 Div. = 10 students
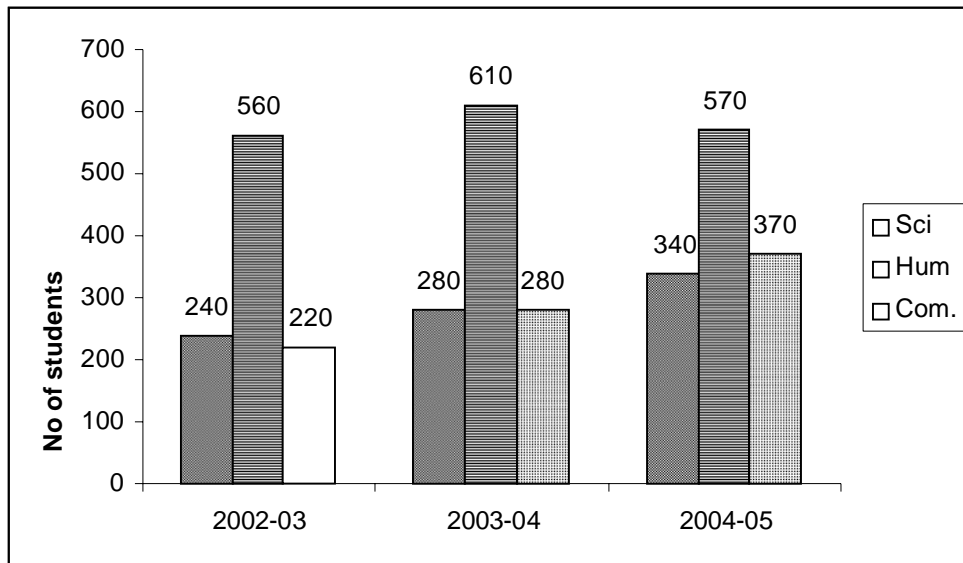


We can present this data by multiple bar diagram in the following manner according torequirements.

Scale x axis 1 cm. = 2 lakhs,on graph 1 Div. = 2. Lakh.
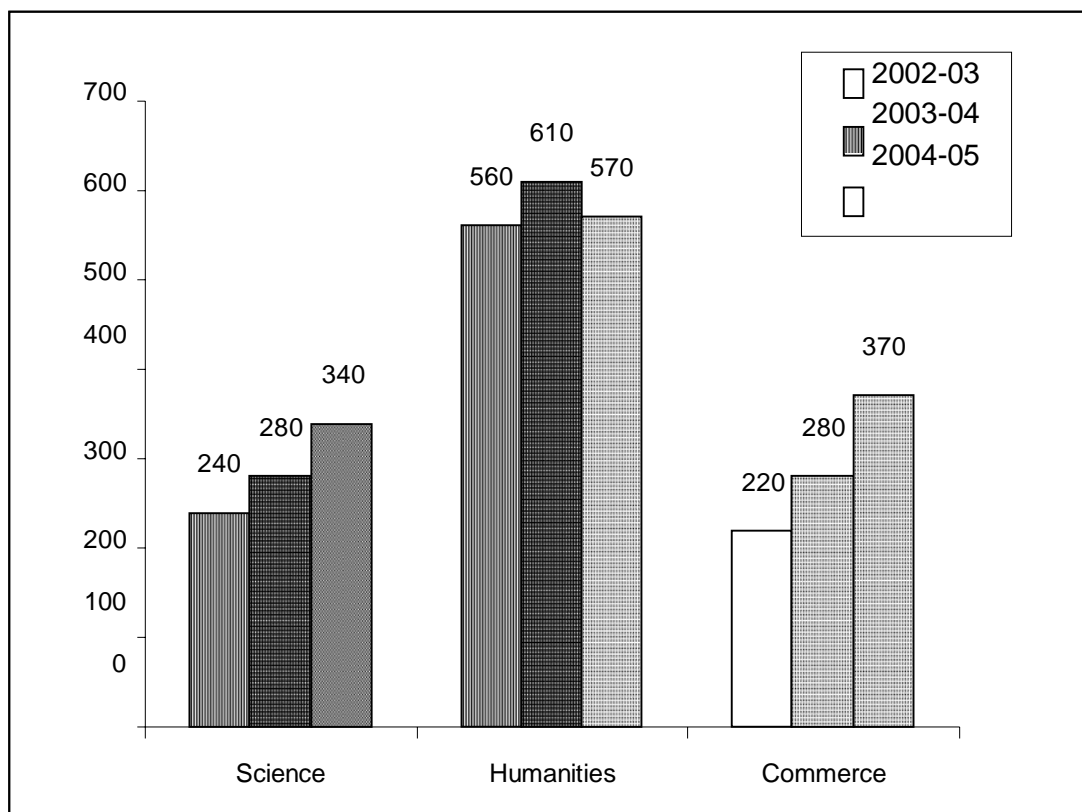
### 4.6.3   Sub-divided Bar Diagram :

| | | Production(Tonnes) | Cumu lativ e | Production (Tonnes) | Cumu lativ e | Production (Tonnes) | Cumu lativ e |
|---|---|---|---|---|---|---|---|
| Wheat | | 8000 | 8000 | 9000 | 9000 | 8500 | 8500 |
| Maiz e | | 4000 | 12000 | 6000 | 15000 | 6000 | 14500 |
| Paddy | | 12000 | 24000 | 11500 | 26500 | 13000 | 27500 |
| | | 24000 | | 26500 | | 27500 | |

The data which is presented by multiple bar diagram can be presented by this diagram.  In this case we add different variables for a period and draw it on a single bar as shown in the following examples. The components must be kept in same order in each bar. This diagram is more efficient if number of components is less i.e. 3 to 5.

**Example 5.**

Production of grains in Punjab is as follows.  Present  the data by a suitable diagram

| Production in Tonnes | Wheat | Maize | Paddy |
|---|---|---|---|
| 2002-03 | 8000 | 4000 | 12000 |
| 2003-04 | 9000 | 6000 | 11500 |
| 2004-05 | 8500 | 6000 | 13000 |

**Solution**

| **2002-03** | **2003-04** | **2004-05** |
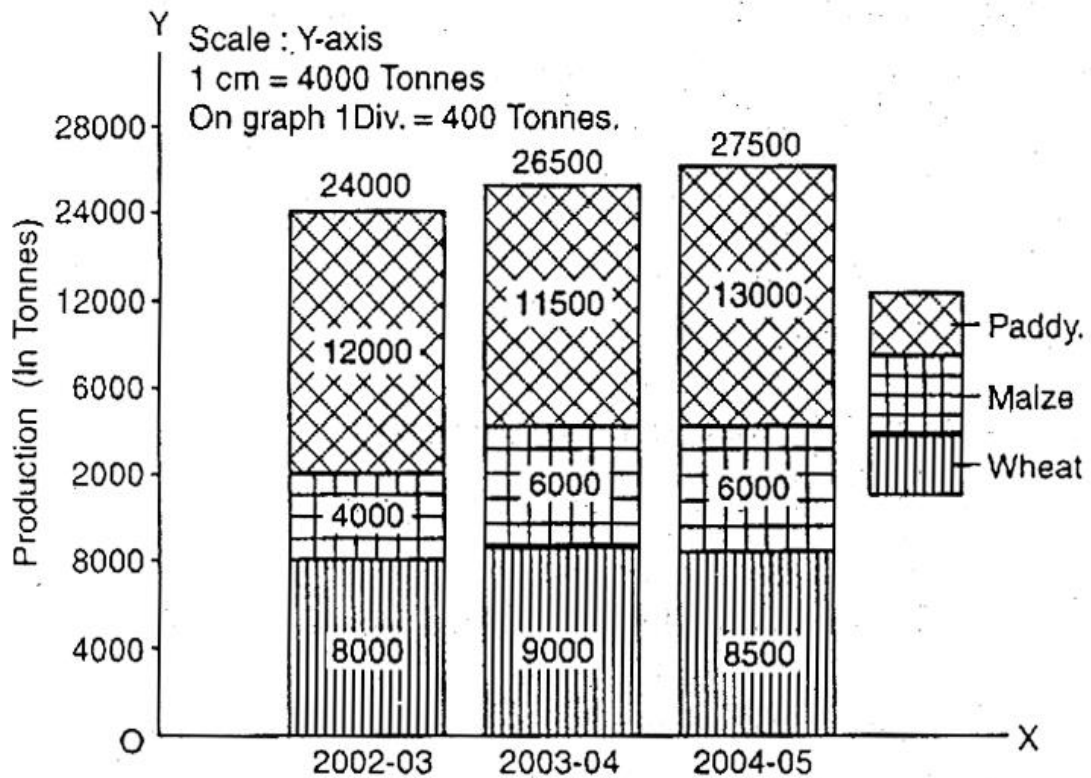
### 4.6.4 Percentage Bar Diagram :

Like sub-divided bar diagram, in this case also data of one particular period or variable is put on single bar, but in terms of percentages, Components are kept in the same order in each bar for easy comparison.

**Example 6.** Present data of Example 4 by percentage bar diagram.

**Solution :**

| Subject | 2002-03 | | | 2003-04 | | | 2004-05 | | |
|---------|------|-------|----------|------|-------|----------|------|-------|----------|
| | No. | Cumu. | % Cumu. | No. | Cumu. | % Cumu. | No. | Cumu. | % Cumu. |
| Science | 240 | 240 | 26 | 280 | 280 | 24 | 340 | 340 | 27 |
| Humanities | 560 | 800 | 78 | 610 | 890 | 76 | 570 | 910 | 71 |
| Commerce | 220 | 1020 | 100 | 280 | 1170 | 100 | 370 | 1280 | 100 |
| | 1020 | | | 1170 | | | 1280 | | |
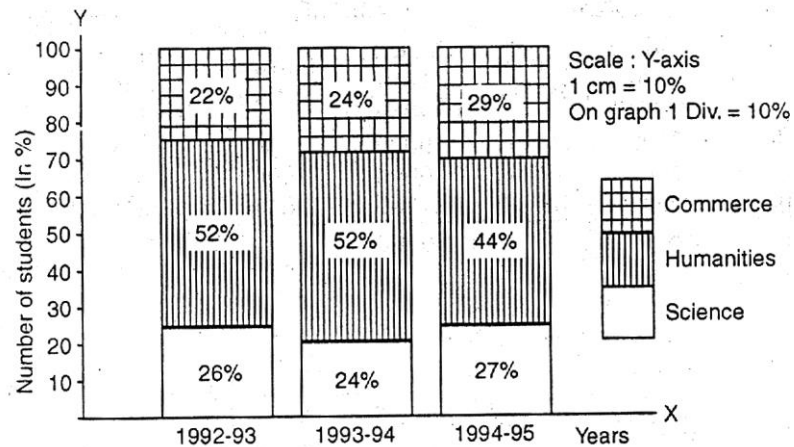
### 4.6.5 Broken Bar Diagram :

This diagram is used when value of some variable is very high or low as compared to others. In this case the bars with bigger terms or items may be shown broken.

**Example 7 :** Present the data given below by suitable diagram

| Year | : | 2002 | 2003 | 2004 | 2005 |
|------|---|------|------|-------|------|
| No. of students | : | 550 | 1000 | 25550 | 700 |

**Solution :**
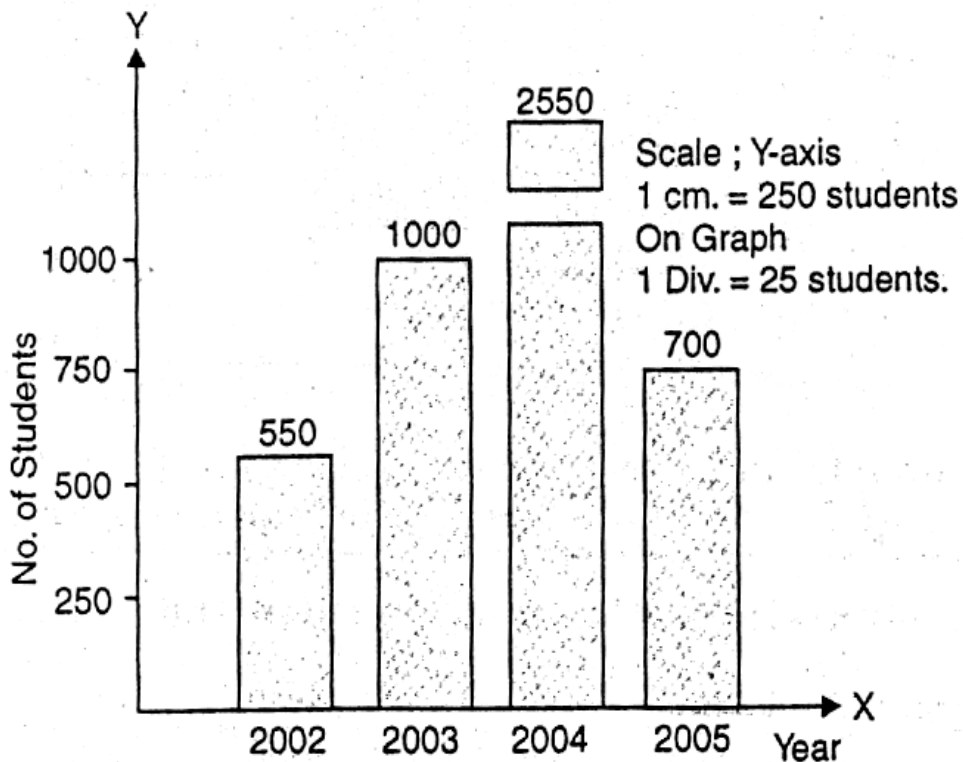
## 4.7 TWO DIMENSIONAL DIAGRAMS :

As in single bars it was mentioned that the width of each bar should be equal for a certain variable or items. But in this case not only the length but the width also is taken proportionately in case of rectangles.

But where the items are represented in square terms we use the technique of squares or circles. These are also known as Area Diagrams.

## 4.7.1 Rectangles :

As mentioned above, not only the length, but the breadth or width of each item is also taken proportionately.

**Example 8.** Present the data given below by rectangle diagram.

| No. of Students | Science | Humanities | Commerce | Total |
|---|---|---|---|---|
| 2002-03 | 170 | 280 | 150 | 600 |
| 2003-04 | 220 | 310 | 270 | 800 |
| 2004-05 | 300 | 360 | 340 | 1000 |

**Solution :**

As the total number of students are in the ratio 3:4:5, we will take the width of bars in this ratio.

Scale Y-axis 1 cm. = 100 studentson graph 1 Div. = 10 students.
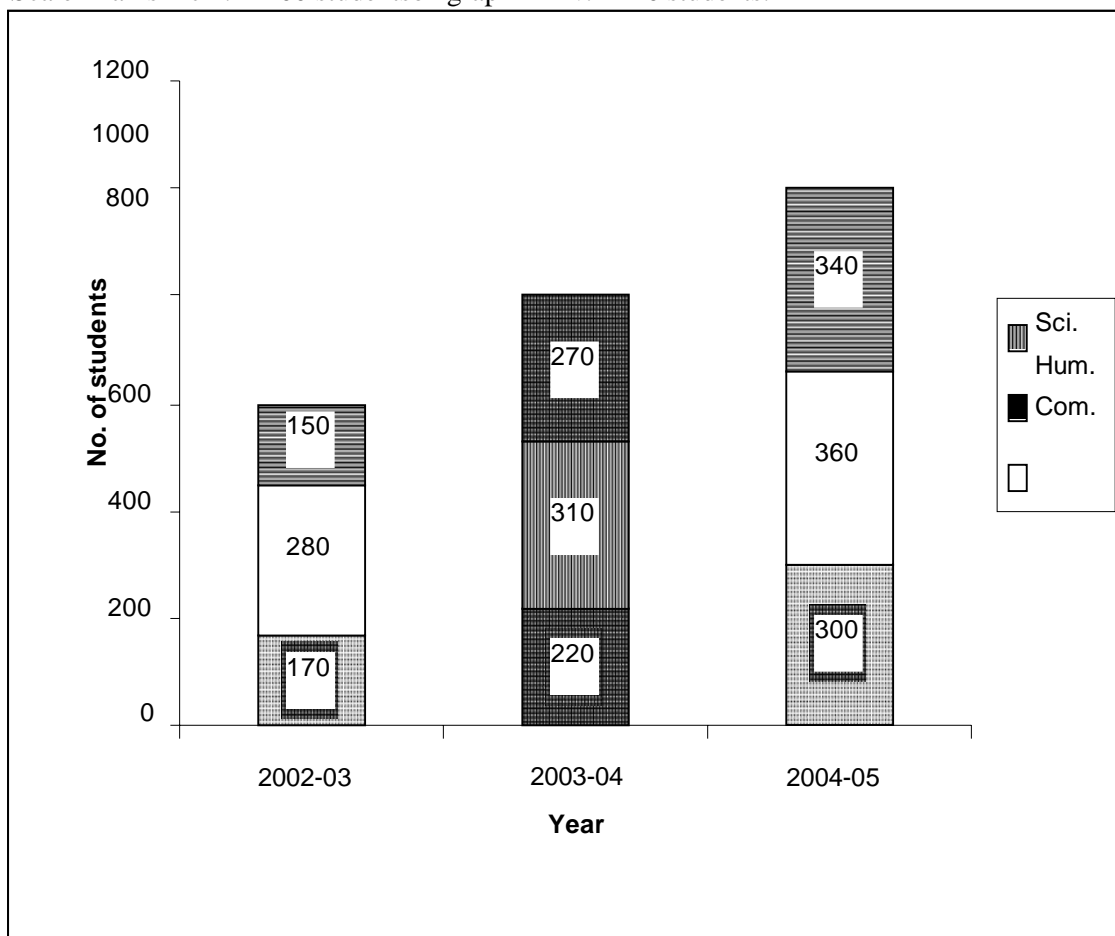
**Example 9.** Income of two families is Rs. 6000 and Rs. 9000 respectively. Present the following data by percentage rectangle diagram.

| Expenditure | Food | Clothing | Fuel | Rent | Education | Miscellaneous |
|---|---|---|---|---|---|---|
| Family A | 3300 | 1080 | 300 | 720 | 480 | 420 |
| Family B | 4500 | 1080 | 450 | 900 | 900 | 720 |

**Solution :**

As the income of two families are in the ratio 2:3 the width of bars should be in the ratio.

| Total Income / Expenditure | Family A | | | Family B | | |
|---|---|---|---|---|---|---|
| | Amount 6000 | % age 100% | Cum. %age | Amount 9000 | % age 100% | Cum. %age |
| Food | 3300 | 55 | 55 | 4500 | 50 | 50 |
| Clothing | 1080 | 18 | 73 | 1080 | 12 | 62 |
| Fuel | 300 | 5 | 78 | 480 | 5 | 67 |
| Rent | 720 | 12 | 90 | 900 | 10 | 77 |
| Education | 480 | 8 | 98 | 900 | 10 | 87 |
| Miscellaneous | 420 | 7 | 105 | 720 | 8 | 95 |
| Total | 6300 | 105 | | 8550 | 95 | |



Scale : 1 cm = 10%
On graph 1 Div. = 10%
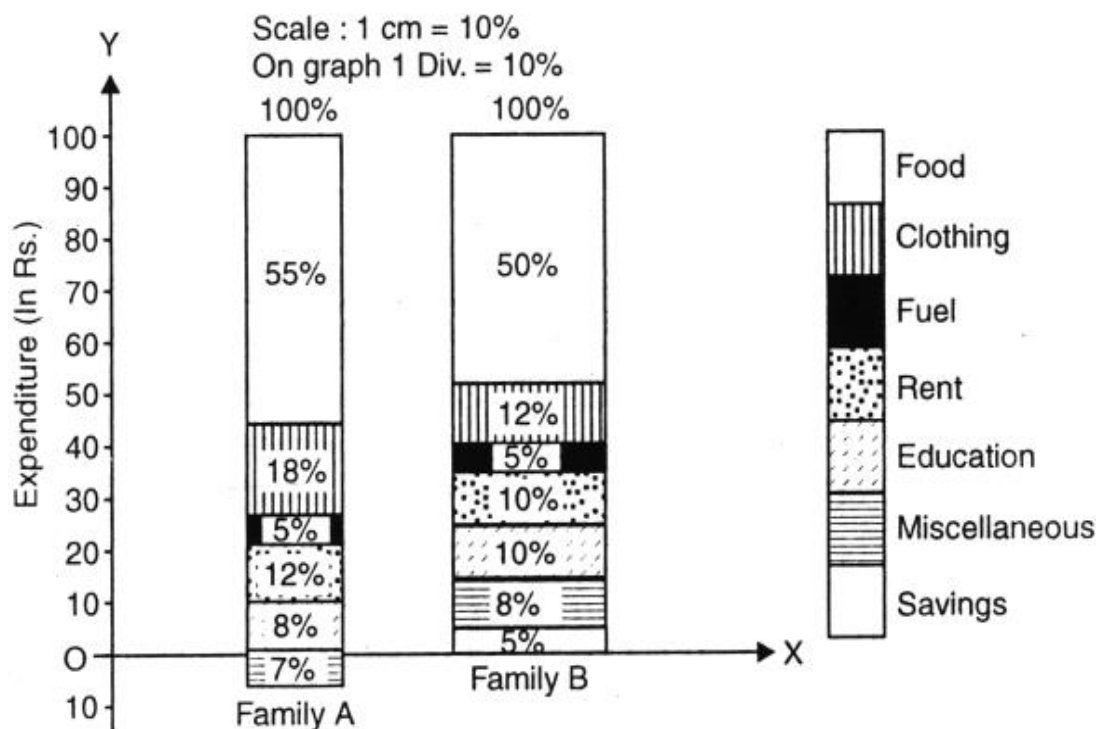
**4.7.2   Squares :**

As told earlier, this technique can be used effectively when given items or terms are squares, preferably having two zeros (00) after every term.

Here we take square root of every item and then divide it by a suitable digit or number so as to get the size reduced to be put into the shape of a square on the given space.

It is also useful technique when difference between the numbers is large.

**Example 10.** Following is the population of some cities in thousands. Present by a suitable

diagram.City                    :          Mumbai    Calcutta    Chennai     Delhi Andhra

Pradesh Population ('00)     :          3600        2500        1600        900     400

**Solution :**

The figures are the perfect squares, hence most suitable diagram will be square or

circle. As side$\sqrt{Area}$          ; in case of square.

So, we take square root of each term.

| City | Population (00) | Square Root | Side (Dividing by 20) |
|---|---|---|---|
| Mumbai | 3600 | 60 | 3.0 cms |
| Calcutta | 2500 | 50 | 2.5 cms |
| Chennai | 1600 | 40 | 2.0 cms |
| Delhi | 900 | 30 | 1.5 cms |
| Andhra Pradesh | 400 | 20 | 1.0 cms |
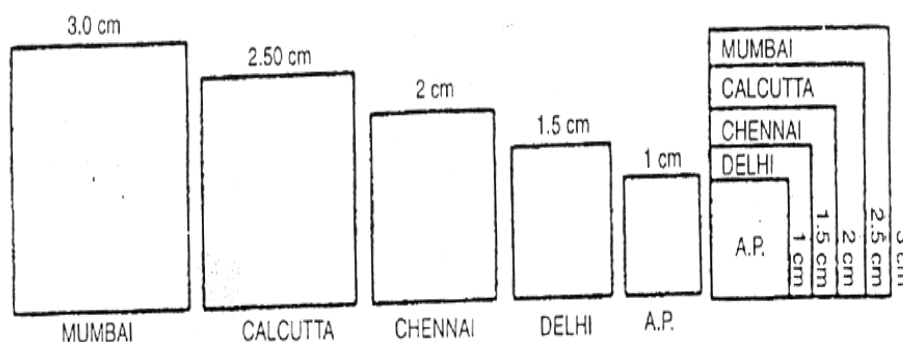


**4.7.3   Sub - divided circular diagram :**

These are also called Pie or Angular Diagrams. We take the total of items and each item is given its proportionate angle taking the total as 360°. In this case we may have to compare in terms of totals

also, if data belongs to two cases.

**Example 11.** Present the the following data through a pie - chart.

| Items | Food | Clothing | Education | Recreation | Misc. |
|---|---|---|---|---|---|
| Expenditure (Rs) : | 5100 | 1200 | 750 | 500 | 75 |

**Solution :**

| Items | Expenditure (Rs.) | Angle |
|---|---|---|
| Food | 5100 | $\dfrac{5100}{7625} \times 360° = 240.72°$ |
| Clothing | 1200 | $\dfrac{1200}{7625} \times 360° = 56.64°$ |
| Education | 750 | $\dfrac{750}{7625} \times 360° = 35.45°$ |
| Recreation | 500 | $\dfrac{500}{7625} \times 360° = 23.65°$ |
| Misc. | 75 | $\dfrac{75}{7625} \times 360° = 3.54°$ |
| Total | 7625 | 360°.00° |

To determine the radius, we take the square - root of 7625, which is 87.32. Divide it by 43.66 to get the radius = 2 cm. Draw a circle with radius = 2 cm and plot the angles obtained in the table. This is called pie-presentation of data.
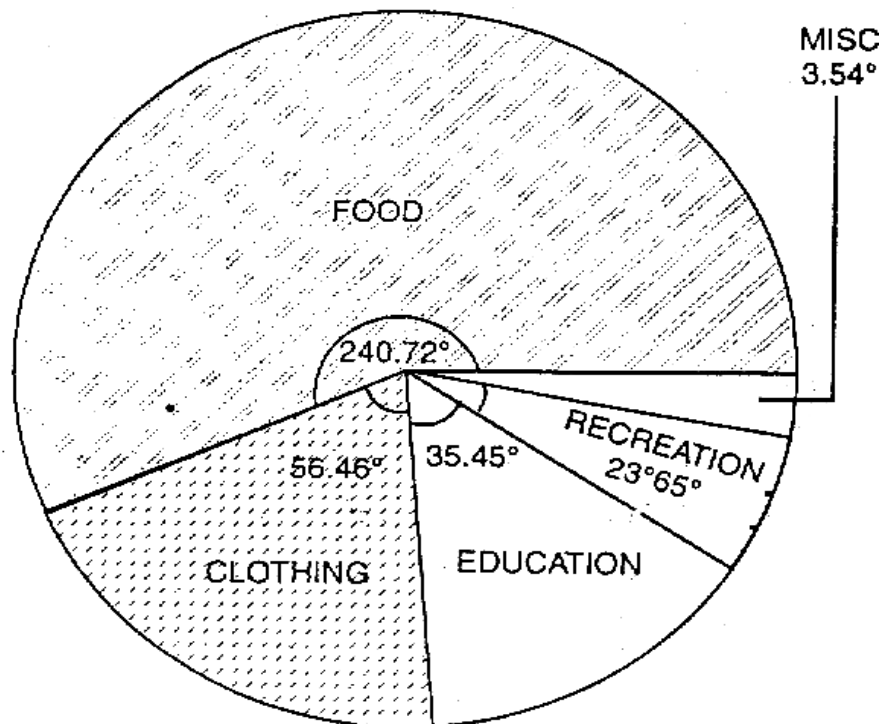
**Example 12.** Prepare Pie Diagram for the following data.

| Items | Food | Clothes | Fuel | Rent | Education | Misc. |
|---|---|---|---|---|---|---|
| Family A: | 1700 | 800 | 400 | 400 | 200 | 100 |
| Family B : | 3300 | 900 | 800 | 600 | 600 | 200 |

**Solution :**

| Item | Family A | | Family B | |
|---|---|---|---|---|
| | **Expenditure** | **Angle** | **Expenditure** | **Angle** |
| Food | 1700 | $\dfrac{1700}{3600} \times 360 = 170°$ | 3300 | $\dfrac{3300}{6400} \times 360 = 180.6°$ |
| Clothes | 800 | $\dfrac{800}{3600} \times 360 = 80°$ | 900 | $\dfrac{900}{6400} \times 360 = 50.6°$ |
| Fuel | 400 | $\dfrac{400}{3600} \times 360 = 40°$ | 800 | $\dfrac{800}{6400} \times 360 = 45.0°$ |
| Rent | 400 | $\dfrac{400}{3600} \times 360 = 40°$ | 600 | $\dfrac{600}{6400} \times 360 = 33.8°$ |
| Education | 200 | $\dfrac{200}{3600} \times 360 = 20°$ | 600 | $\dfrac{600}{6400} \times 360 = 33.8°$ |
| Miscellaneous | 100 | $\dfrac{100}{3600} \times 360 = 10°$ | 200 | $\dfrac{200}{6400} \times 360 = 11.2°$ |
| Total | 3600 | 360° | 6400 | 360° |

Taking square roots of 3600 and 6400 we get, 60 and 80. We can divide it by a common denominator 40, to get their radii as 1.5 and 2.0 cms.
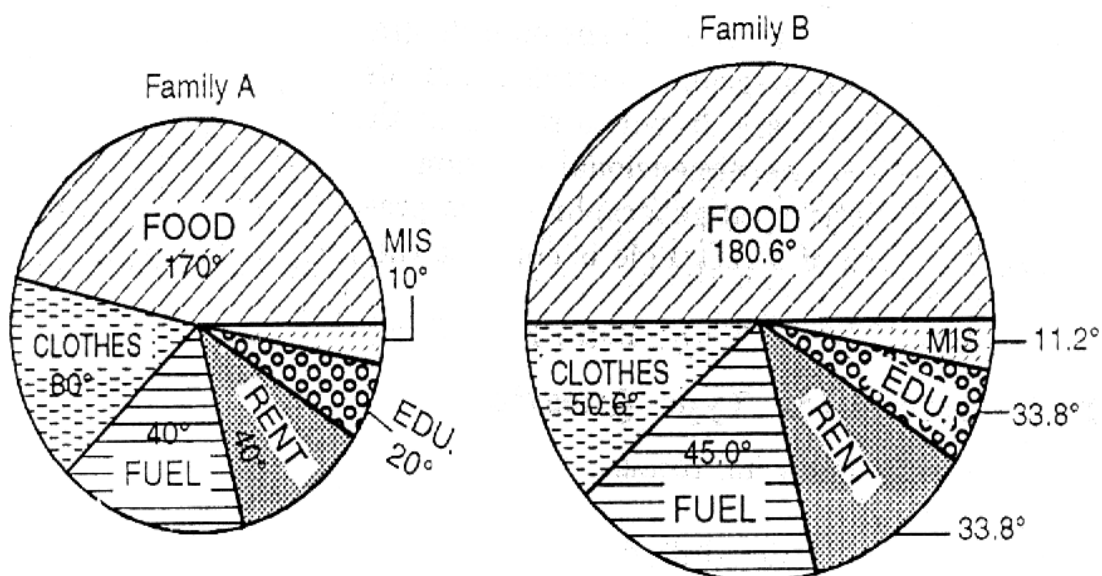


**Note:** If there is only single case, we may take any length of radius to suit our space.

**GRAPHS :**

## 4.8  INTRODUCTION :

Diagrams can present the data in an attractive style but still there is a method more reliable than this. Diagrams are often used for publicity purposes but are not of much use in statistical analysis. Hence graphic presentation is more effective and meaningful.

According to A.L. Boddington,

"The wandering of a line is more powerful in its effect on the mind than a tabulated statement; it shows what is happening and what is likely to take place, just as quickly as the eye is capable of working."

## 4.9  USES OR MERITS OR IMPORTANCE OF GRAPHS :

1  It is more effective than diagrams.

2  It is economical in terms of money, times and energy.

3  It gives us the picture in condensed form.

4  It  is free from mathematical calculations.

5  It is most suitable for comparison.

6  It is helpful in forecasting.

7  It is also helpful in statistical analysis. We can determine Median and Mode by this method.

## 4.10  RULES OR GUIDELINES WHILE PREPARING A GRAPH :

1.  It must have proper heading.

2.  Scale must be provided along with graph,

3.  False base line may be used according to need.

4.  If required footnotes may be given.

5.  While choosing scale, size of the space must be kept in view,

6.  If possible Y-axis should be about 50% more than X-axis,

7.  Independent variables should be taken on X-axis and dependent variable on Y axis.

8.  In time series graph, time should be shown on X-axis and other variable on Y-axis.

9.  In frequency distribution, take value of variable on X-axis and frequency on Y-axis.

## 4.11  CONSTITUTION OF GRAPH PAPER :

Graphs are drawn on a special type of paper known as graph paper. Graph papers are divided in small equal squares $\frac{1}{10}$ or $\frac{1}{10}$ cm.

For the construction of graph, two straight lines, are drawn which cut each other at $90^r$. The horizontal line is called 'X'-axis and is usually denoted by X'OX. The vertical line is

called Y-axis and is usually denoted by Y'OY. The point where they cut each other is known as 'Origin'.

This- origin divides the graph paper in four parts. These parts are known as quadrants.

Zero value is taken on the point of origin for both lines. Positive values of X are taken towards fight side on horizontal line and of Y towards upper side on vertical line. Negative values of X are taken towards the left side on horizontal line and of Y towards the lower side on vertical line.

**Positive and Negative Values :**
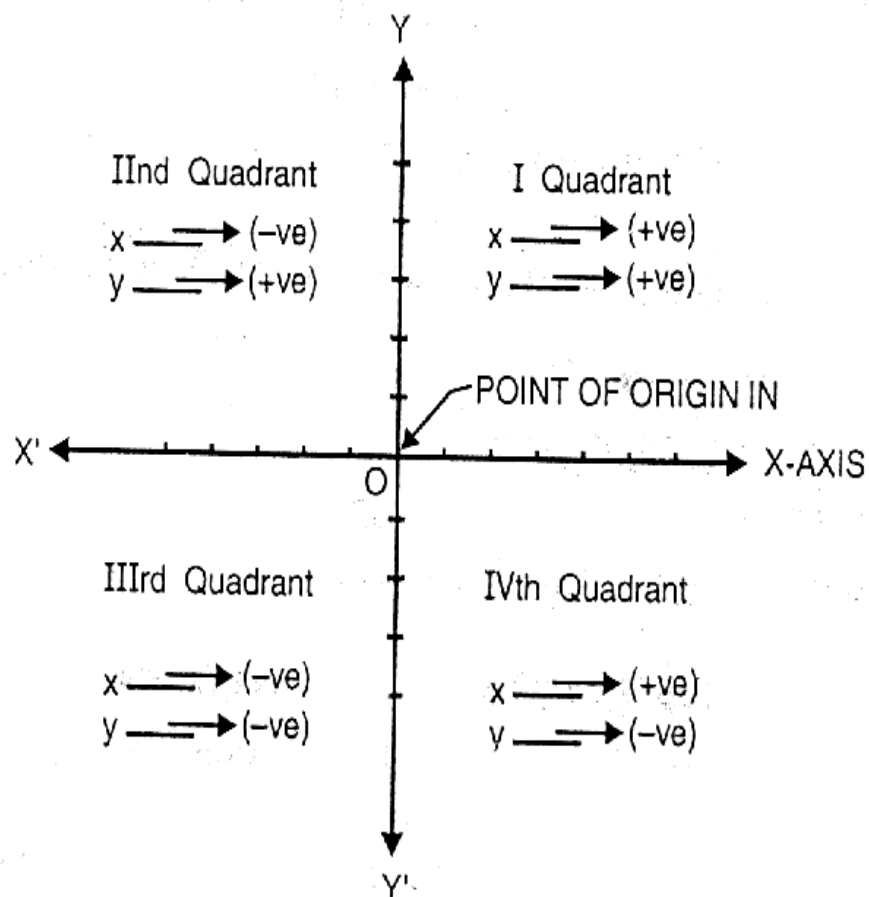


As shown in above diagram in first quadrant X and Y both have positive values. In second, X is negative and Y positive. In third, X and Y both are negative and in fourth Quadrant X is positive and Y negative.

X-axis is also known as abscissa and Y-axis as Ordinate.

### 4.12 CHOICE OF SCALE :

The scale indicates the unit of a variable that a fixed length of axis would represent. Scale may be different for both the axes.. It should be taken in such a way so as to accommodate whole of the data *on* a given graph paper in a lucid and attractive style.

**4.13 FALSE BASE LINE :**

Sometimes it is difficult to take zero at origin and proceed for the graph as is in the followingexample :

| Year | : | 2001 | 2002 | 2003 | 2004 |
|---|---|---|---|---|---|
| No. of students | : | 10320 | 10860 | 11400 | 11200 |

If we start with zero in this case, first fifty main divisions will remain empty, without any use.

In will make the graph look clumsy. In such cases we use false base lines as shown below.
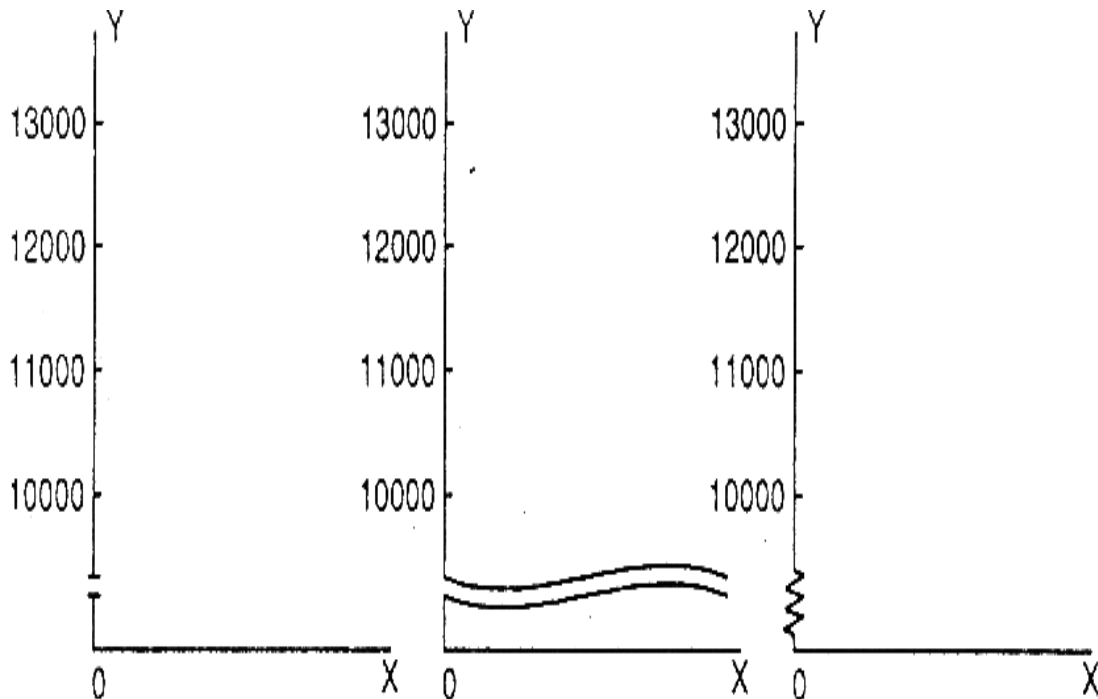


If required we can take flase base line on x - axis also.

**4.14    TYPES OF GRAPHS :**

There are two types of graphs.

Time series Graphs or Historigrams

Frequency Distribution Graphs.

**4.14.1  Time series graphs or Historigrams :**

In this type of graphs, time is taken along X-axis and the other variables along Y-axis. The number of variables on Y-axis may be one or more than one. These are known as One Variable,Two Variables or Three Variables graphs.

**Example 1.** Present following data on a graph paper. (Single variables )

| Year | : | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
|---|---|---|---|---|---|---|---|---|---|---|
| Pay/month | : | 370 | 395 | 405 | 419 | 439 | 456 | 472 | 472 | 473 |

**Solution :**

We will take here false base line.



Scale on Graph
Y-axis 1 Div. = 2 Units
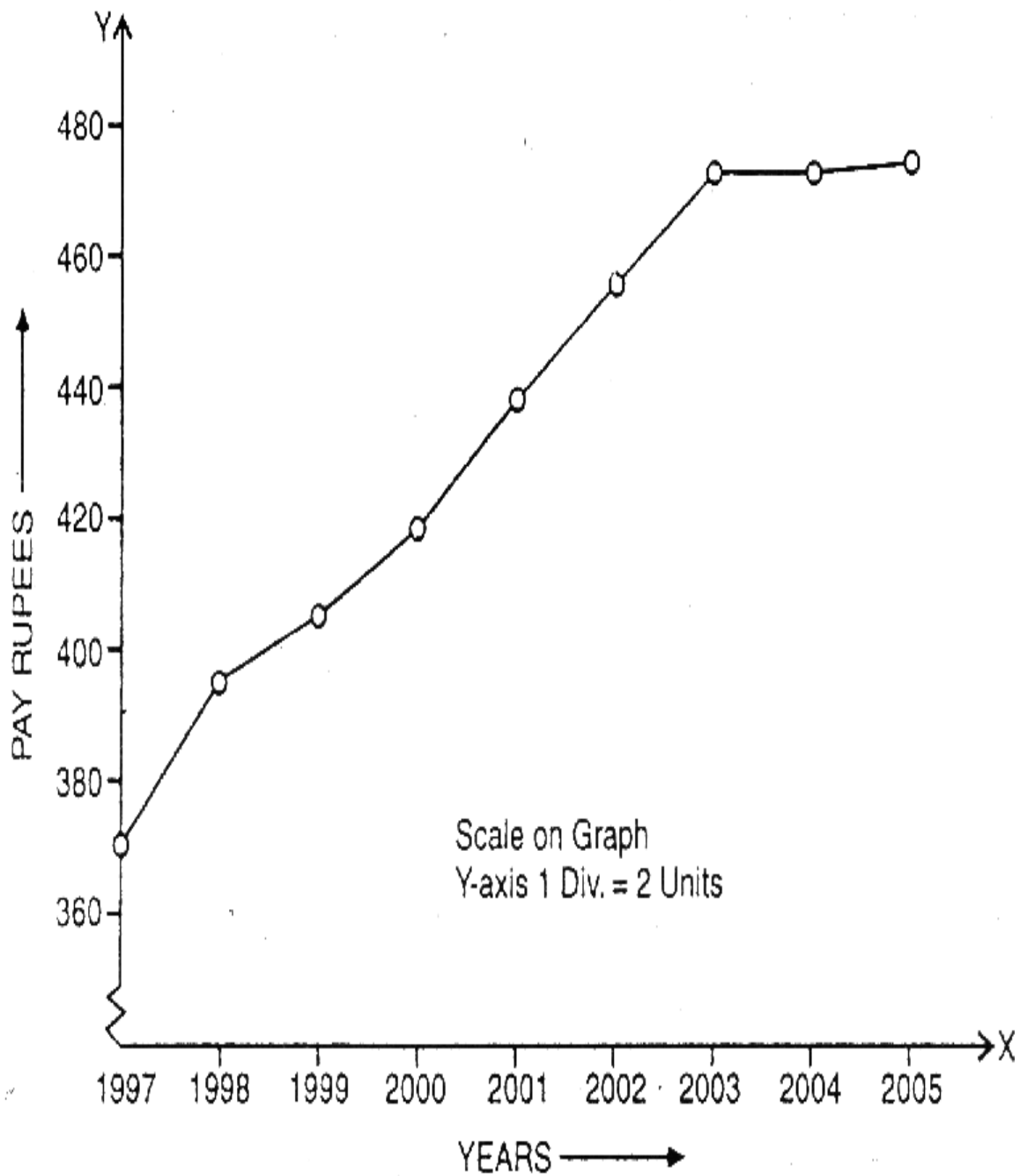
**Example 2.** Present following data on a graph paper. ( Two variables )

| Year | : | 20001-02 | 2002-03 | 2003 -04 | 2004-05 |
|---|---|---|---|---|---|
| Value of imports (in Rs. Crores) | | 2000 | 2500 | 3500 | 4500 |
| Value of Exports (in Rs. Crores) | | 1700 | 2800 | 4400 | 4900 |

**Solution :**

We will  take here false base line.
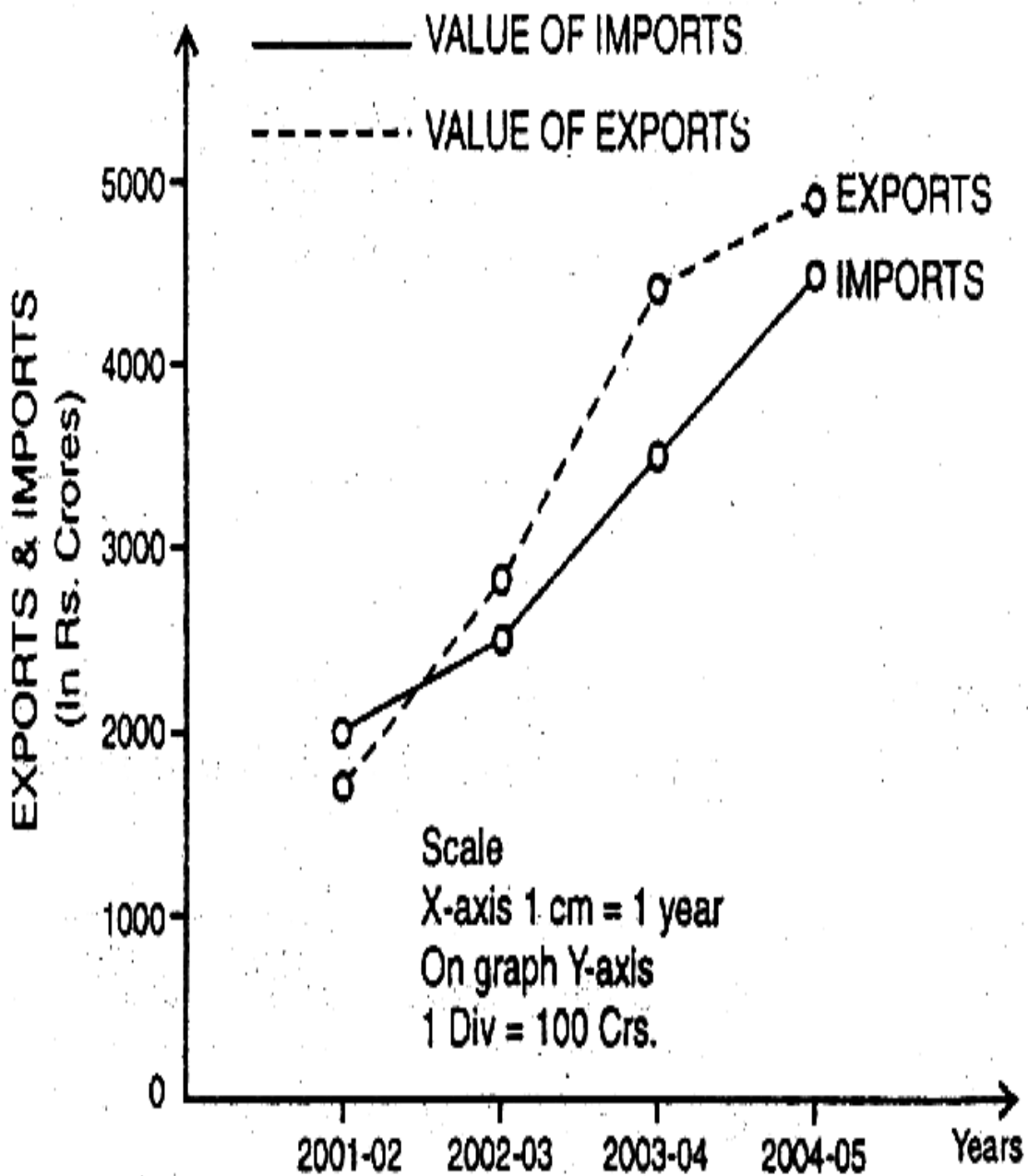
**4.14.2  Graphs of frequency distribution :**

A  frequency  distribution  can  be graphically presented in the following manner :

1.  Histogram

2.  Frequency Polygon

3.  Smoothed frequency curves

4.  Cumulative Frequency Curves or Ogives

**4.14.2.1  Histogram :**

The term histogram should not be confused with the term histogram which represents time charts. Histogram or column diagram is the best way. of presenting graphically a simple frequency distribution. The classes are marked along the X-axis and by taking class-interval as the base rectangles are erected with heights proportional to the respective classes. Frequencies are measured along the Y-axis. With equal class intervals, all rectangles will have equal base and the area of each rectangle will be proportional to the frequency in that class. In case of unequal class intervals' the width of the rectangles will change and the heights of rectangles shall be proportional to the density of the frequency or the adjusted frequencies.

**Construction of Histogram with equal class intervals**

**Example 8.** Prepare a Histogram from the following data.

| Marks | : | 325-350 | 350-375 | 375-400 | 400-425 | 425-450 |
|---|---|---|---|---|---|---|
| No. of Students | : | 30 | 45 | 75 | 60 | 35 |

**Solution :**

Here we take false base line for OX, as smallest term is 325. We take scales for both the axes. For X-axis, 1 Div. $= \dfrac{25}{10} = 2.5$ marks and on Y-axis ; 1 Div. $= \dfrac{10}{10} = 1$ student.
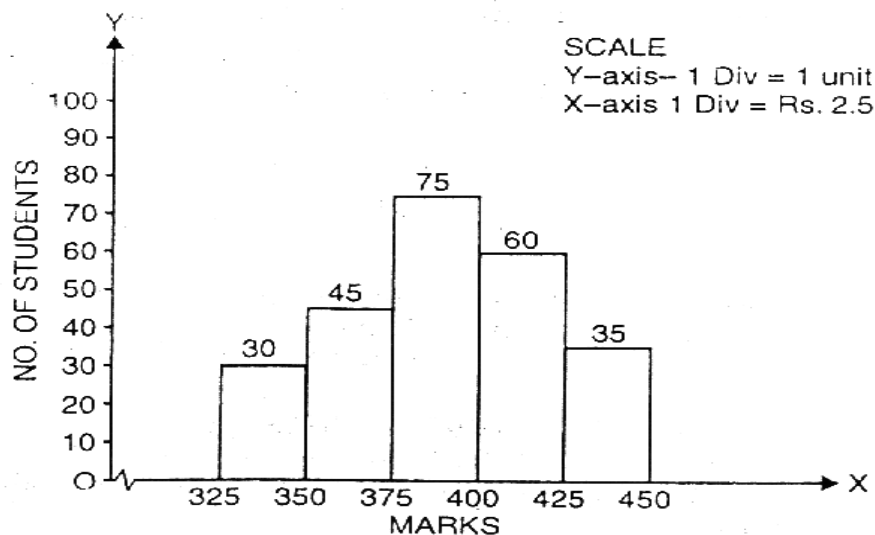
**Construction of Histograms with Unequal class-Intervals :**

When class intervals are unequal, it is necessary to make adjustment for varying magnitude of class intervals by determining frequency densities. First of all we should decide the class- interval in terms of which the frequency density is to be calculated. The most common interval is generally taken. Then we convert the frequencies of all those classes which have a larger or smaller class-interval to frequencies in terms of class-interval already decided.

**Example 9.** Prepare a Histogram from the following data.

| Marks | : | 10-15 | 15-20 | 20-25 | 25-30 | 30-40 | 40-60 | 60-80 |
|---|---|---|---|---|---|---|---|---|
| No. of Students | : | 7 | 19 | 27 | 15 | 12 | 12 | 8 |

**Solution :**

Since the class intervals are unequal, frequencies, must be adjusted. Here the most common smallest class interval is 5. We convert the interval 12 size 30-40 in two intervals and the frequency is divided by 2 *i.e.* $\frac{12}{2} = 6$ Similarly 40-60 is divided into 4 and the frequency is divided by

i.e. $\frac{12}{4} = 3$ and 60 -80 also into 4 and its frequency is also divided by 4 i.e. $\frac{8}{4} = 2$.

| Class Intervals (Marks) | 10-15 | 15-20 | 20-25 | 25-30 | 30-35 | 35-40 | 40-45 | 45-50 | 50-55 | 55-60 | 60-65 | 65-70 | 70-75 | 75-80 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequency (No. of Students) | 7 | 19 | 27 | 15 | 6 | 6 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 2 |

The Histogram is drawn on its basis is given below

**4.14.2.2 Frequency Polygon :**

A frequency polygon is a curve representing a frequency distribution. If we join the middle points of the tops of the adjacent rectangles of the histogram, a frequency polygen is obtained. Here both the ends of the polygon are extended to the base line so that the area under the polygon is equal to the area under the histogram. The value of mode can easily be found by forming a frequency perpendicular from the apex of the polygon to the X-axis.
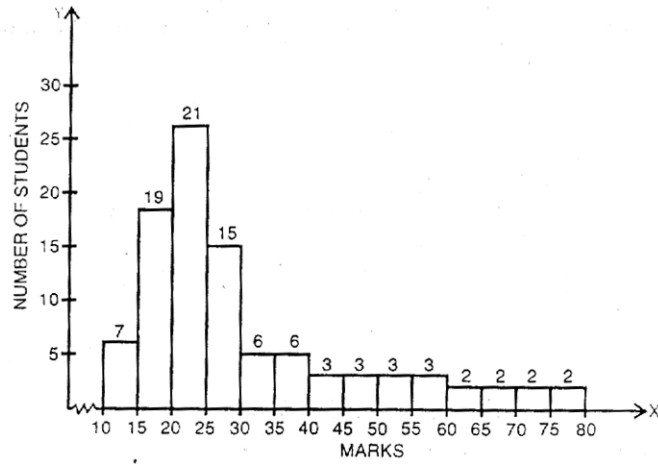
**Example 10 :** Prepare a histogram and frequency polygon from the following data.

| X | f |
|---|---|
| 0-10 | 13 |
| 10-20 | 17 |
| 20-30 | 15 |
| 30-40 | 13 |
| 40-60 | 10 |

**Solution :**
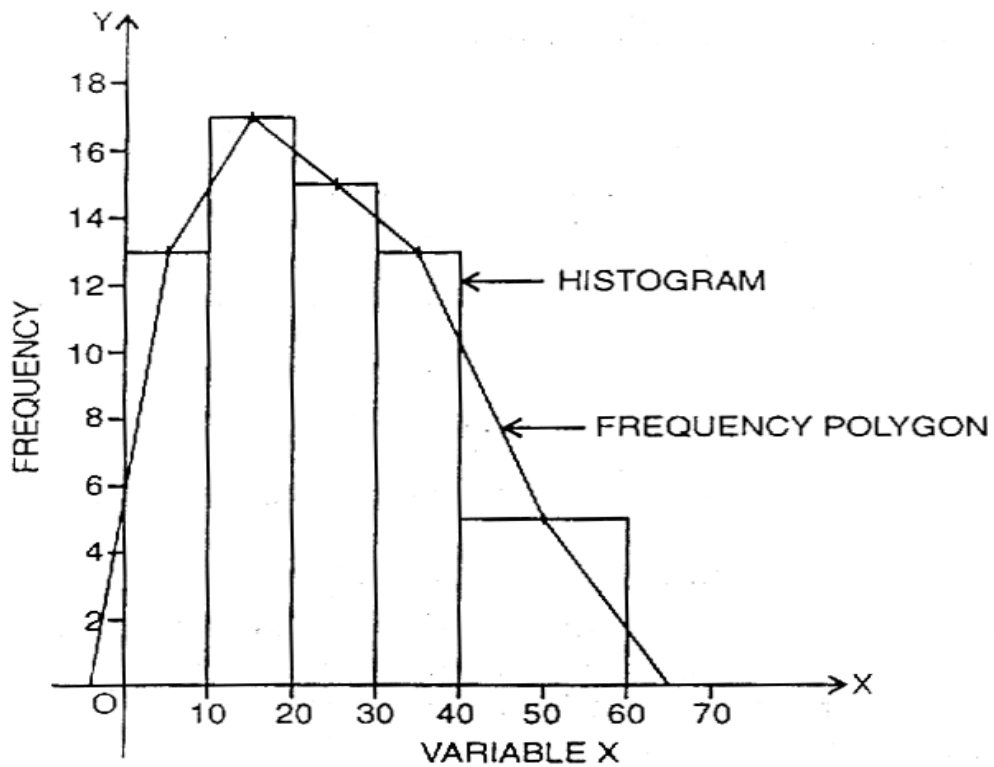
The class 40 -60 is presented with frequency 5 (10/2 = 5)

**Example 11 :** From the following data, determine the modal value graphically and verify the result by actual calculation.
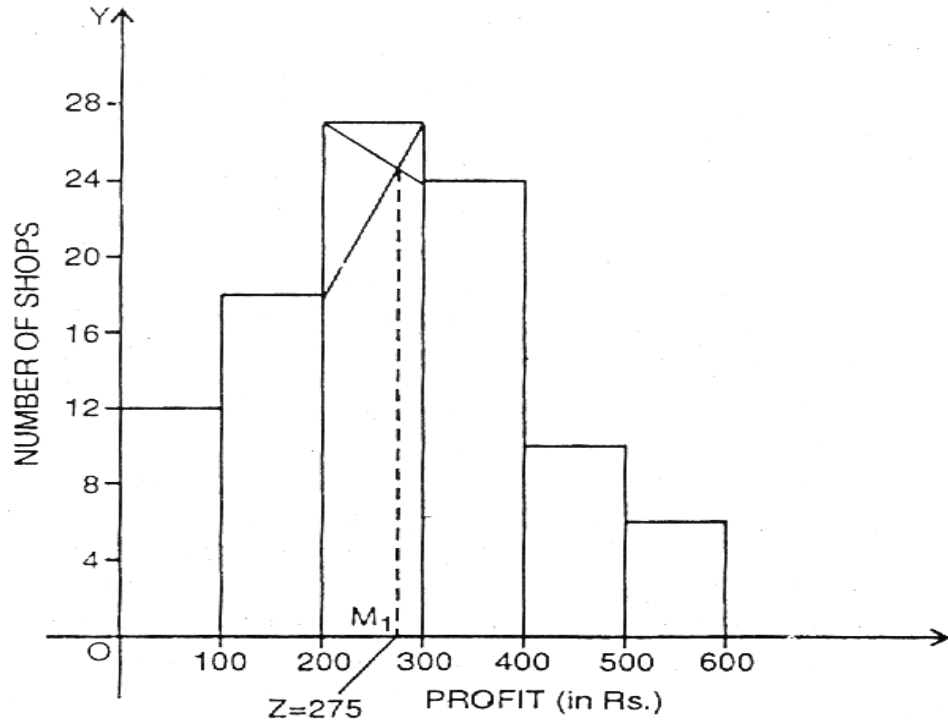


### 4.14.2.3 Smoothed Frequency Curve :

This curve emerges when the points of a frequency polygon are joined by free hand smoothed curves and not. by straight lines. The area of the frequency polygon is equal to that of the histogram. This curve should be based on samples and only continuous series should be smoothed.

**Example 12 .** Draw a histogram, frequency Polygram and frequency curve representing the following data.

| Daily Wages ( in Rs): | 0-20 | 20-40 | 40-60 | 60-80 | 80-100 |
|---|---|---|---|---|---|
| No. of Students : | 20 | 50 | 90 | 38 | 15 |

**Solution :**

Here in addition to the construction of histogram, frequency polygon, frequency curve is drawn by smoothing the corners of the frequency polygon as shown below :
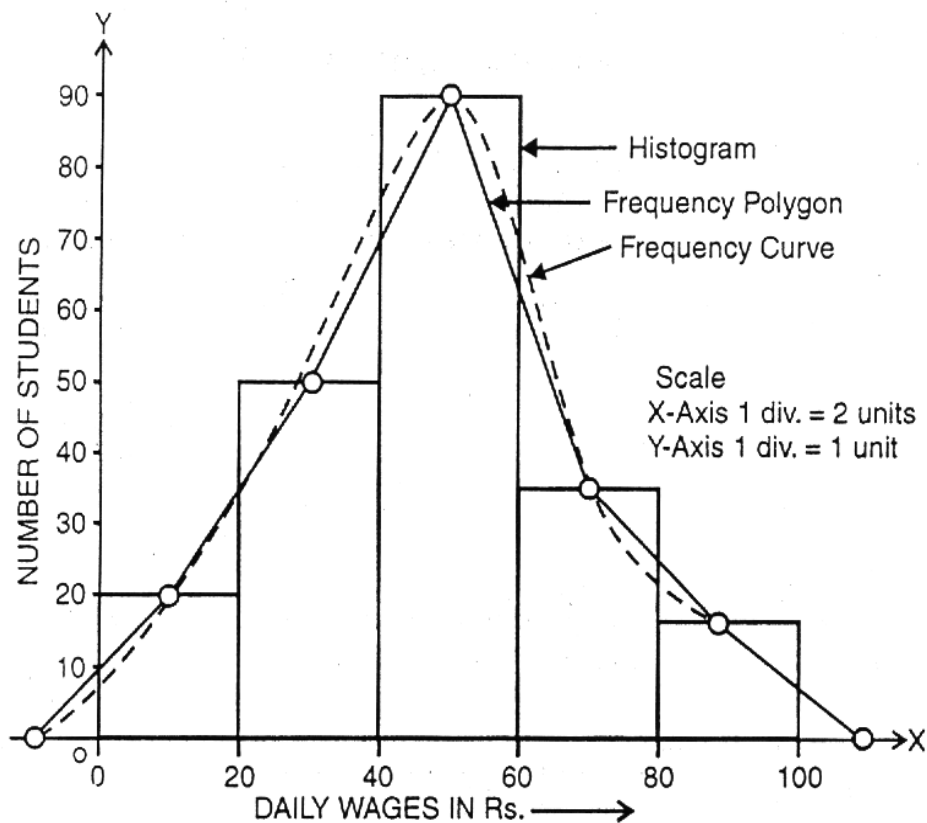
## 4.14.2.4 Cumulative Frequency Curves :

Sometimes it is necessary to know the number of items whose values are more than or less than a certain amount this case we have to change the form of frequency distribution from a simple to a cumulative distribution. The graphic representation of cumulative frequency distribution is called the cumulative frequency curve or Ogive.

There are two methods of drawing ogive -

The less than method and

the 'more than method.

If we want to know the number of items that are 'less than' a particular size, the cumulation will start from the least to the greatest size and the series will be called 'less than' cumulative frequency distribution. When we want to know the number of items whose sizes are 'more than' a particular size, cumulation will commence from the greatest to the least and the series thus obtained shall be termed as 'more than' cumulative frequency distribution. Ogives - are used to- determine the number 'or percentage of cases above or below a certain value. Ogives are also used to compare two or more frequency distributions. Ogives can also be used to determine graphically the values of median, quartile, deciles etc.

**Example 13 .** Draw the two gives from the following data and locate Median.

| Class Interval | 100-200 | 200-300 | 300-400 | 400-500 | 500-600 | 600-700 |
|---|---|---|---|---|---|---|
| Frequency : | 12 | 18 | 30 | 42 | 68 | 78 |

**Solution :**

The cumulation frequency distributions are given bellow :

| Class Interval | Cf | Class Interval | Cf |
|---|---|---|---|
| Less than 100 | 0 | More than 100 | 240 |
| " 200 | 12 | " 200 | 228 |
| " 300 | 30 | " 300 | 210 |
| " 400 | 60 | " 400 | 180 |
| " 500 | 102 | " 500 | 138 |
| " 600 | 162 | " 600 | 78 |
| " 700 | 240 | 700 | 0 |

Table showing calculation of Median



| Class Interval | f | Cf |
|---|---|---|
| 100–200 | 12 | 12 |
| 200–300 | 18 | 30 |
| 300–400 | 30 | 60 |
| 400–500 | 42 | 102 |
| 500–600 | 60 | 162 |
| 600–700 | 78 | 240 |

Median Class

$$= \frac{N}{2} \text{ th item}$$

$$= \frac{240}{2} = \text{120th item}$$
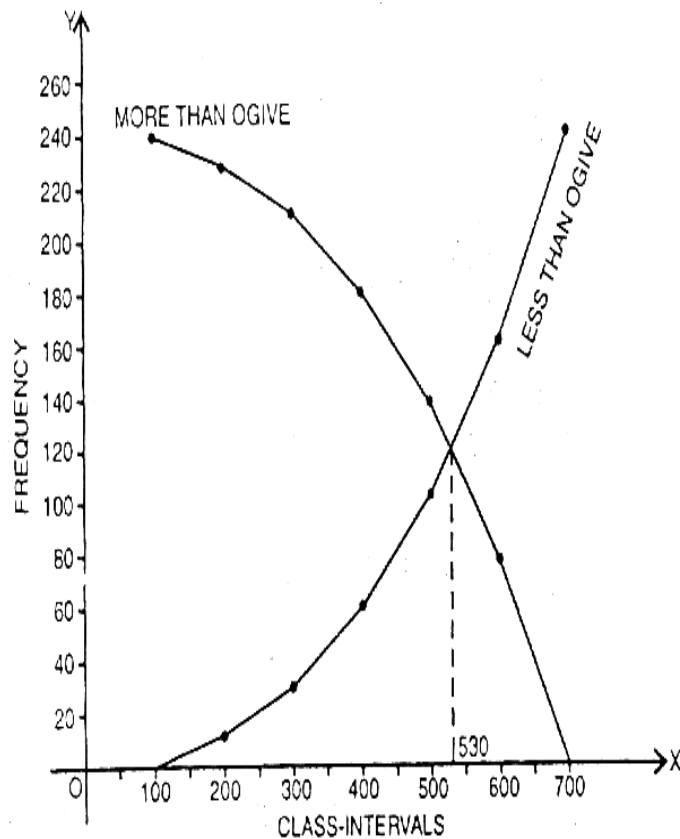
So median class is 500 -600

$$M = L + \frac{\frac{N}{2} - Cf}{f} \times i = 500 +$$

## 4.15   SUMMARY :

Graphs are simple and attractive. Both can give condensed form to data and help to compare the variables. Even a layman or an illiterate person can easily understand the diagrams & graphs.

## 4.16   QUESTIONS :

1.   What do you mean by a 'Diagram' ?

2.   What are the limitations of diagrams ?

3.   Define various types of diagrams.

4.   Define One or Single dimension diagram.

5.   Narrate merits of one dimension diagram.

6.   Define line diagram. How will you draw it ?

7.   What is simple bar diagram ? How will you draw it ?

8.   What is multiple  diagram ? When is it used ?

9.   What is two dimensional diagram ? Define its various types.

10.  When is a rectangle or a square or a circle is used to present a data ?

11.  Explain the necessity of diagrams in statistics.

12.  Explain the need and usefulness of diagrammatic representation of statistical data. What are the different types  of diagrams you know ?

13.  What is a pie diagram ? Draw a pie-Diagram with imaginary figures of children, adolescents, middle age and old age people in a particular place.

14.  Describe the steps involved in the construction of a pie diagram.

15.  What do you mean by a graph  ?

16.  How to choose scale for a graph ?

                        Or

     What points should be taken on the base while selecting  a scale for a graph ?

17.  What do you mean by  false base line ? When is it used and How ?

18.   Define various  types' of graphs.

19.   Explain time series graph

             Or
     What is histogram ?

20.   Define various types of frequency distribution graphs.

21.  What is a histogram ? How to draw it ?

22. Define frequency polygon. How to draw it ?

23. Define Ogive. What are its types ? How to construct all these ?

24.    (a)What is the difference between Diagrams and Graphs ?

       (b)Name the graph that are used to locate mode and, median respectively.

25. Define uses or merits or importance of Graphs.

26. What steps or guidelines should be followed to prepare a good graph

## 4.17 EXERCISES :

Draw line diagram to present the following data

(a)  Class         :M.Com.      M.Sc.  M.A   M.B.B.S      B.E.

     No. of students :   220     180     340      80      120

(b)

     Country:          U.S.A       U.K  Japan   India Pakistan France

Per capita income (in 000):        32        22    28      4              20

2. Present the following data by a bar diagram.

| Country | Production of sugar capitaincome (in 000) |
|---|---|
| Cuba | 32 |
| Australia | 30 |
| India | 20 |
| Japan | 5 |
| Jawa | 1 |
| Egypt | 1 |

2. Present the following data by multiple

diagram.(a)

| Course | No.of students | | |
|---|---|---|---|
| | **2002-03** | **2003-04** | **2004-05** |
| M.A | 420 | 320 | 380 |
| M.Sc. | 200 | 240 | 360 |
| M.Com. | 140 | 300 | 480 |

(b)

| Dept. | No. of students | | |
|---|---|---|---|
| | **2002-03** | **2003-04** | **2004-05** |
| Arts | 600 | 550 | 500 |
| Science | 400 | 500 | 600 |
| Commerce | 200 | 250 | 300 |

3. Draw sub-divided bar diagrams for 3 (a) and (b)

4. Draw a suitable diagram to present the following information.

| | | Selling Price | Qt. Sold | Wages | Material | Misc. | Totals |
|---|---|---|---|---|---|---|---|
| Factory x | : | 400 | 20 | 3200 | 2400 | 1600 | 7200 |
| Factory y | : | 640 | 30 | 6000 | 6000 | 9000 | 21000 |

( Hint : Preferably draw percentage bar diagram)

5. Present following data on a graph paper. (Single

variables )(a)

| Year | : 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
|---|---|---|---|---|---|---|
| No.of students | : 210 | 380 | 410 | 540 | 430 | 360 |

b)

| Year | : 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
|---|---|---|---|---|---|---|

( Hint : For (b), take false base )

6. Value of imports and exports is given; Draw graph.

| Year | : | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
|---|---|---|---|---|---|---|---|
| Imports | : | 1080 | 1120 | 1240 | 1360 | 1040 | 1260 |
| Exports (In Core Rs.) : | | 640 | 120 | 980 | 1240 | 1340 | 1120 |

**T. Nageswara Rao**

**Vice - Principal**

**Hindu College, Guntur.**

# LESSON - 5
# AVERAGES – I

**OBJECTIVES :**

After studying this lesson you should be able to understand the following.

1. Measure of Central Value

2. Objectives and features of Averages

3. Types of Averages

4. Merits and Demerits of Averages

5. Arithmetic Mean and its Calculations

**STRUCTURE OF THE LESSON :**

## 5.1    MEANING OF AVERAGE :

The main objective of statistical analysis is to arrive at one single value which represents the whole series. This value is called central value or an average. The value of average has a tendency towards centralization. That means it lies in the middle of the data. It is the reason that averages are sometimes called measures of central tendency.

The word average is very commonly used in day-to-day conversation. For example, we often

talk of average boy in a long, average height etc. It is defined by different statisticians.

**a.**      **According to Ya-Lun – Chou :**

"An average is a typical value in the sense that it is sometimes employed to represent all the individual values in a series or of a variable".

**b.**      **According to 'Croxton and Cowden :**

"An average value is a single value within the range of the data that is used to represent all of the value in the series. Since an average is somewhere within the range of the data, it is something called a measure of central value".

## 5.2    OBJECTIVES OF AVERAGE :

There are two main objectives of the study of averages:

1. To get single value that describes the characteristic of the entire group.

2. Measures of Central value condenses the mass of data in one single value, enable us to get remember the date easily.

3. Measures of Central values, by reducing the mass of data to one single figure, enable compari sons to be made. For example, we can compare the percentage results of the students of differentcolleges in a certain examination.

4. Averages are useful in decision making.

## 5.3    REQUISITES OF A GOOD AVERAGE :

A typical average should possess the following essentials or ideals to be a good average.

1. **It should be easy to understand:** Since statistical methods are designed to simplify complexity, it is desirable that an average be such that can be readily understood; other-wise, its use is bound to be very limited.

2. **It should be simple to compute** : An average should be simple to compute so that itcan be used widely.

3. **It should be based on all the items**: The average should depend upon each and every item of the series so that if any of the items is dropped the average itself is altered.

4. **It should no be unduly affected by extreme observations** : If one or two very smallor very large items unduly affect the averages i.e. either increase its value or reduce its value, the average cannot be reall typical of the entire series.

5. **It should be rigidly defined:** The average should be properly defined so that it has one and only one interpretation. It should preferably be defined by an algebraic formula so that if different people compute the average from the same figures they all get the same answer.

6. **It should be capable of further algebraic treatment :** We should prefer to have an average that could be used for further statistical commutations so that its utility is enhanced.

## 5.4 TYPES OF AVERAGES :

The following are the important types of averages.

     1.   Arithmetic Mean

2. Median

3. Mode

4. Geometric Mean

5. Harmonic Mean

## 5.5  ARITHMETIC MEAN :

The most popular and widely used measure of representing the entire data by one value is what most laymen call an 'average' and what the statisticians call the arithmetic mean.  Its value is obtained by adding together all the items and be dividing this total by the number of items. Arithmetic mean is two types.

Arithmetic Mean

Simple Arithmetic Mean
Mean

Weighted Arithmetic
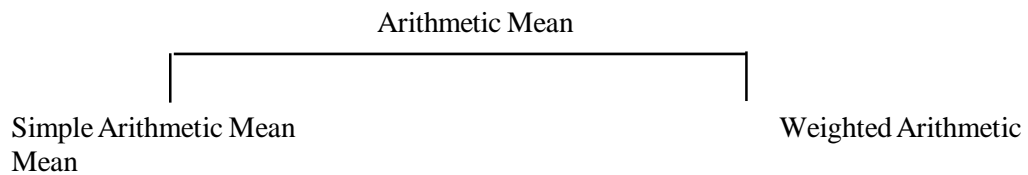
### 5.5.1  Arithmetic Mean - Individual Series :

The process of computing mean in case of individual observations is very simple. Add together the various values of the variable and divide the total by the number of items.

**Direct Method :**

$$\text{Arithmetic Mean} = \frac{\Sigma x}{N}$$

$\Sigma x$ = Total

of Values N

= Number

of items.

**Illustration 1** : Calculate Arithmetic Mean from the following

| data.S.No. | : | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Daily Wages | : | 75 | 43 | 57 | 21 | 49 | 39 | 80 | 12 | 95 | 59 |

| Sno. | Daily Wages Rs. |
|---|---|
| 1 | 75 |
| 2 | 43 |
| 3 | 57 |
| 4 | 21 |
| 5 | 49 |
| 6 | 39 |
| 7 | 80 |
| 8 | 12 |
| 9 | 95 |
| 10 | 59 |
| **10** | **530** |

$$\text{Arithmetic Mean} = \frac{\Sigma x}{N}$$

$\Sigma x = 530$

$N = 10$

$$\text{Arithmetic Mean} = \frac{530}{10}$$

$$= 53$$

**Illustration 2 :**

Calculate Arithmetic Mean from the following Values.

Values 43     48     68     57     31     60     37     48     78

| Values |
| --- |
| 43 |
| 48 |
| 65 |
| 57 |
| 31 |
| 60 |
| 37 |
| 48 |
| 78 |
| **526** |

$$\text{Arithmetic Mean} = \frac{\Sigma x}{N}$$

$\Sigma x = 526$

$N = 10$

$$\text{Arithmetic Mean} = \frac{526}{10}$$

Arithmetic Mean = 52.6

 = 52.6

**Shortcut Method** : The arithmetic mean can be calculated by using what is known as an arbitrary origin, when deviations are taken from the arbitrary origin, the formula for calculating arithmetic mean is -

$$\text{Arithmetic Mean} = X \pm \frac{\Sigma f dx}{N}$$

X = Assumed Mean

N = Number of items

Σfdx = Summation of multiples of deviations with their corresponding frequencies.

**Illustration 3**

Calculation Arithmetic Mean from the following information.

Values :         27    24    29    25    26    23    34    12    19    30    32

**Solution**

| Values X | 27 | 24 | 29 | 25 | 26 | 23 | 34 | 12 | 19 | 30 | 32 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| dx | +2 (27-25) | -1 (24-25) | +4 (29-25) | 0 (25-25) | +1 (26-25) | -2 (23-25) | +9 (34-25) | -13 (12-25) | -6 (19-25) | +5 (30-25) | +7 (32-25) | +28-22 =+6 |

$$\text{Arithmetic Mean} = X + \frac{\Sigma fdx}{N}$$

X = 25

N = 11

Σfdx = +6

$$\text{Arithmetic Mean} = 25 + \frac{6}{11}$$

Arithmetic Mean = 25 + .54 = 25.54

Arithmetic Mean = 25.54

**Illustration 4** : From the following Marks calculate Arithmetic Mean.

Marks :43    48    65    57    31    60    37    48    78    59

**Solution**

| Marks | 43 | 48 | 65 | 57 | 31 | 60 | 37 | 48 | 78 | 59 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| dx | -14 | -9 | +8 | 0 | +26 | +3 | -20 | -9 | +21 | +2 | **-78+34 = - 44** |

$$\text{Arithmetic Mean} = X + \frac{\Sigma fdx}{N}$$

X = 57

N = 10

Σfdx = -44

Arithmetic Mean $= 57 + \dfrac{-44}{10}$

Arithmetic Mean $= 57 - 4.4 = 52.6$

Arithmetic Mean $= 52.6$

**Illustration 5 :** From the following data calculate Arithmetic

Mean.Family :A    B    C    D    E    F    G    H    I    J

   Salary Rs.:85    70    10    75    500    8    42    250    40    36

Solution :

| Family | A | B | C | D | E | F | G | H | I | J | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Salary (x) (Rs.) | 85 | 70 | 10 | 75 | 500 | 8 | 42 | 250 | 40 | 36 | |
| dx | +10 | -5 | -65 | 0 | +425 | -67 | -33 | +175 | -35 | -39 | -244 +610 =+366 |

Arithmetic Mean $= X + \dfrac{\Sigma fdx}{N}$

    $X = 75$

    $N = 10$

    $\Sigma fdx = +366$

    Arithmetic Mean $= 75 + \dfrac{366}{10}$

    Arithmetic Mean $= 75 + 36.6$

    Arithmetic Mean $= 116.6$

## 5.5.2 Arithmetic Mean - Discrete Series :

In Discrete Series Arithmetic Mean may be computed by applying either Direct method or Short-cut method.

**Direct Method :**

Arithmetic Mean $= \dfrac{\Sigma xf}{N}$

$\Sigma xf$ = Multiply the frequency of each row with the variable and obtain the total of xfDivide the total obtained by the number of observations. (i.e. total frequency)

**Illustration 6 :**

From the following data. Calculate Arithmetic Mean of 40 workers.

| Wages (Rs.)  : | 3 | 5 | 8 | 10 | 12 | 15 |
|---|---|---|---|---|---|---|
| No. of Workers: | 4 | 10 | 12 | 8 | 4 | 2 |

**Solution :**

| Wages (X) (Rs.) | 3 | 5 | 8 | 10 | 12 | 15 | |
|---|---|---|---|---|---|---|---|
| No. of Workers (f) | 4 | 10 | 12 | 8 | 4 | 2 | **40** |
| xf | 12 | 50 | 96 | 80 | 48 | 30 | **316** |

$$\text{Arithmetic Mean} = \frac{\Sigma xf}{N}$$

$$\Sigma xf = 316$$
$$N = 40$$

$$\text{Arithmetic Mean} = \frac{316}{40}$$

$$\text{Arithmetic Mean} = 7.9$$

**Illustration 7** : From the following data calculate Arithmetic mean of 100 employees.

| Salary (Rs.)   : | 40 | 60 | 80 | 100 | 120 | 140 | 160 | 180 | 200 | |
|---|---|---|---|---|---|---|---|---|---|---|
| No.of   : | 5 | 7 | 10 | 15 | 20 | 25 | 9 | 6 | 3 | |
| Employees | | | | | | | | | | |
| **Solution** Salary (X) (Rs.) | 40 | 60 | 80 | 100 | 120 | 140 | 160 | 180 | 200 | |

| No. of Employees | 5 | 7 | 10 | 15 | 20 | 25 | 9 | 6 | 3 | **100** |
|---|---|---|---|---|---|---|---|---|---|---|
| xf | 200 | 420 | 800 | 1500 | 2400 | 3500 | 1440 | 1080 | 600 | **11,940** |

Arithmetic Mean $= \dfrac{\Sigma xf}{N}$

$\Sigma xf = 11{,}940$

$N = 100$

Arithmetic Mean $= \dfrac{11940}{100}$

Arithmetic Mean $= 119.4$

## Short-cut Method :

$$\overline{X} = X + \dfrac{\Sigma fdx}{N}$$

X = assumed mean

N = total of the frequency

$\Sigma fdx$ = Sum of multiples of deviations with their frequency

## Illustration 8 :

Calculate Arithmetic Mean from the following data.

| Values | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Frequency | 7 | 11 | 16 | 17 | 26 | 31 | 11 | 1 | 1 |

**Solution**

| Values X | Frequency (f) | dx | f X dx fdx |
|---|---|---|---|
| 1 | 7 | -4 | -28 |
| 2 | 11 | -3 | -33 |
| 3 | 16 | -2 | -32 |
| 4 | 17 | -1 | -17 |
| 5 | 26 | 0 | 0 |
| 6 | 31 | +1 | +31 |
| 7 | 11 | +2 | +22 |
| 8 | 1 | +3 | +3 |
| 9 | 1 | +4 | +4 |

| | | | -110+60 = -50 |
|---|---|---|---|

$$\text{Arithmetic Mean} = X + \frac{\Sigma fdx}{N}$$

X = 5

N = 121

$\Sigma fdx$ = -50

$$\text{AM}(X) \quad = 5 + \frac{-50}{121} = 5 + (-0.413)$$

AM = 4.587

**Illustration 9 :** Following is the data of 735 families. Calculate average number of children per families.

| No. of | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Children : | 108 | 154 | 126 | 95 | 62 | 45 | 20 | 11 | 6 | 5 |

0 No. of

families :

96

| No. of Children (X) | No. of families (f) | dx | fdx |
|---|---|---|---|
| 0 | 96 | -6 | -576 |
| 1 | 108 | -5 | -540 |
| 2 | 154 | -4 | -616 |
| 3 | 126 | -3 | - 1378 |
| 4 | 95 | -2 | -190 |
| 5 | 62 | -1 | - 62 |
| 6 | 45 | 0 | 0 |
| 7 | 20 | +1 | +20 |
| 8 | 11 | +2 | +22 |
| 9 | 6 | +3 | +18 |
| 10 | 5 | +4 | +20 |
| 11 | 5 | +5 | +25 |
| 12 | 1 | +6 | +6 |
| 13 | 1 | +7 | +7 |
| | **735** | | **-2362 + 118= -2244** |

$$\text{Arithmetic Mean} = X + \frac{\Sigma fdx}{N}$$

X = 6

N = 735

$\Sigma fdx = -2244$

$$\text{AM (X)} = 6 + \frac{-2244}{735} = 6 + (-3.05)$$

AM = 2.95

## Illustration 10 :

From the following data calculate Average Mark.

| Marks | : | 4 | 5 | 6 | 7 | 8 | 9 |
| No. of Students | : | 8 | 10 | 9 | 6 | 4 | 3 |

**Solution**

| Marks (X) | No. of students (f) | dx | fdx |
|---|---|---|---|
| 4 | 8 | -2 | -16 |
| 5 | 10 | -1 | -10 |
| 6 | 9 | 0 | 0 |
| 7 | 6 | +1 | +6 |
| 8 | 4 | +2 | +8 |
| 9 | 3 | +3 | +9 |
| | **40** | | **+23-26 = - 3** |

$$\text{Arithmetic Mean} = X + \frac{\Sigma fdx}{N}$$

X = 6

N = 40

$\Sigma fdx = -3$

$$\text{AM (X)} = 6 + \frac{-3}{40} = 6 + (-0.075)$$

AM = 5.925

## Step Deviation Method :

In the step deviation method the only additional point is that in order to simplify calculations we take a common factor from the data and multiply the result by the common factor.

$$\overline{X} = X + \frac{\Sigma fDx}{N} \times C$$

X = assumed mean

N = total frequency

C = Common factor

ΣfDx = Sum of multiplies of Dx with frequency

## Illustration 11 :

From the following information calculate Arithmetic Mean by using Step deviation method.

| Wages | 40 | 60 | 80 | 100 | 120 | 140 | 160 | 180 | 200 |
|---|---|---|---|---|---|---|---|---|---|
| No. of Workers | 5 | 7 | 10 | 15 | 20 | 25 | 9 | 6 | 3 |

**Solution**

| Wages (X) | No. of Workers (f) | dx | DX (20) | fDX (Dx X f) |
|---|---|---|---|---|
| 40 | 5 | -60 | -3 | -15 |
| 60 | 7 | -40 | -2 | -14 |
| 80 | 10 | -20 | -1 | -10 |
| 100 | 15 | 0 | 0 | 0 |
| 120 | 20 | +20 | +1 | +20 |
| 140 | 25 | +40 | +2 | +50 |
| 160 | 9 | +60 | +3 | +27 |
| 180 | 6 | +80 | +4 | +24 |
| 200 | 3 | +100 | +5 | +15 |
| | **100** | | | **-39+136 +97** |

$$\overline{X} = X \pm \frac{\Sigma fDx}{N} \times C$$

X = 100

N = 100

C = 20

ΣfDx = +97

$$\overline{X} = 100 + \frac{97}{100} \times 20 = 100 + 0.97 \times 20$$

$$= 100 + 19.4 = 119.4$$

$$\overline{X} = 119.4$$

**Illustration 12 :** From the following data calculate Arithmetic Mean.

| Marks | 5 | 15 | 25 | 35 | 45 | 55 | 65 | 75 |
|---|---|---|---|---|---|---|---|---|
| No. of Students | 2 | 18 | 30 | 45 | 26 | 20 | 6 |  |

**Solution :**

| Marks (X) | No. of Students (f) | dx | DX (10) | fDX (Dx X f) |
|---|---|---|---|---|
| 5 | 2 | -30 | -3 | -6 |
| 15 | 18 | -20 | -2 | -36 |
| 25 | 30 | -10 | -1 | -30 |
| 35 | 45 | 0 | 0 | 0 |
| 45 | 26 | +10 | +1 | 26 |
| 55 | 20 | +20 | +2 | 40 |
| 65 | 6 | +30 | +3 | 18 |
| 75 | 3 | +40 | +4 | 12 |
|  | **150** |  |  | **-72+96 =+24** |

$$X = \overline{X} + \frac{\Sigma fDx}{N} \times C$$

$X = 35$

$N = 150$

$C = 10$

$\Sigma fDx = +24$

$\overline{X} = 35 + \dfrac{24}{150} \times 10 = 100 + 0.97 \times 20$

$= 35 + 0.16 \times 10 = 35 + 1.6 = 36.6$

$\overline{X} = 36.6$

## 5.5.3 Arithmetic Mean - Continuous Series

In continuous series, arithmetic mean may be computed by applying any of the following methods.

**Direct Method :**
     Short-cut Method
     Step Deviation Method

**Direct Method :**

$$\text{Arithmetic Mean} = \frac{\Sigma fx}{N}$$

$\Sigma fx$ = Sum of multiples of variables with

frequenciesN = Number of variables.

**Illustration 13 :** From the following data calculate Average Profit.

| Profit per shop Rs. in Lakhs | No. of shops |
|---|---|
| 0-10 | 12 |
| 10-20 | 18 |
| 20-30 | 27 |
| 30-40 | 20 |
| 40-50 | 17 |
| 50-60 | 6 |

**Solution**

| Profit per shop Rs. in Lakhs(X) | No. of shops(f) | Mid Point of X | xf |
|---|---|---|---|
| 0-10 | 12 | 5 | 60 |
| 10-20 | 18 | 15 | 270 |
| 20-30 | 27 | 25 | 675 |
| 30-40 | 20 | 35 | 700 |
| 40-50 | 17 | 45 | 765 |
| 50-60 | 6 | 55 | 330 |
| | **100** | | **2800** |

$$\text{Arithmetic Mean} = \frac{\Sigma fx}{N}$$

$\Sigma fx = 2800$

$N = 100$

$$\text{Arithmetic Mean} = \frac{2800}{100}$$

$= 28$ Lakhs

$\overline{X}$ = 28

**Illustration 14 :**

From the following wages of 40 workers calculate average wage.

| Wage Rs. | No. of Workers |
|---|---|
| 130-140 | 3 |
| 140-150 | 15 |
| 150-160 | 10 |
| 160-170 | 8 |
| 170-180 | 3 |
| 180-190 | 1 |

**Solution :**

| Wage Rs.(X) | No. of Workers(f) | Mid Point of X | xf |
|---|---|---|---|
| 130-140 | 3 | 135 | 405 |
| 140-150 | 15 | 145 | 2175 |
| 150-160 | 10 | 155 | 1550 |
| 160-170 | 8 | 165 | 1320 |
| 170-180 | 3 | 175 | 525 |
| 180-190 | 1 | 185 | 185 |
| | **40** | | **6160** |

$$\text{Arithmetic Mean} = \frac{\Sigma fx}{N}$$

$\Sigma fx = 6160$

$N = 40$

$$\text{Arithmetic Mean} = \frac{6160}{40} = 154$$

$\overline{X} = 154$

**Short Cut Method :**

$$\overline{X} = X + \frac{\Sigma fdx}{}$$

N

X = assumed mean

$\Sigma fdx$ = sum of multiples of dx with frequenciesN = total of the frequency

**Illustration 15 :** Calculate Arithmetic Mean from the following data.

| Values | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|--------|------|-------|-------|-------|-------|-------|-------|
| Frequencies | 7 | 11 | 17 | 21 | 14 | 6 | 4 |

| Values | Frequency | Mid Point of X | dx | fdx |
|--------|-----------|----------------|-----|-----|
| 0-10 | 7 | 5 | -30 | -210 |
| 10-20 | 11 | 15 | -20 | -220 |
| 20-30 | 17 | 25 | -10 | -170 |
| 30-40 | 21 | <u>35</u> | 0 | 0 |
| 40-50 | 14 | 45 | +10 | +140 |
| 50-60 | 6 | 55 | +20 | +120 |
| 60-70 | 4 | 65 | +30 | +120 |
|  | **80** |  |  | **-600+380** <br> **-220** |

$$\overline{X} = X + \frac{\Sigma fdx}{N}$$

$X = 35$

$\Sigma fdx = -220$

$N = 80$

$$\overline{X} = 35 + \frac{-220}{80}$$

= 35 + - 2.75

Arithmeic Mean = 32.25

## Illustration 16 :

From the following data calculate Average Wage

| Wages Rs. | 130-140 | 140-150 | 150-160 | 160-170 | 170-180 | 180-190 |
|-----------|---------|---------|---------|---------|---------|---------|
| No. of Workers | 3 | 15 | 10 | 8 | 3 | 1 |

**Solution :**

| Wages Rs. | No. of Wokers | Mid Point of X | dx | fdx |
|---|---|---|---|---|
| 130-140 | 3 | 135 | -20 | -60 |
| 140-150 | 15 | 145 | -10 | -150 |
| 150-160 | 10 | 155 | 0 | 0 |
| 160-170 | 8 | 165 | +10 | +80 |
| 170-180 | 3 | 175 | +20 | +60 |
| 180-190 | 1 | 185 | +30 | +30 |
| | **40** | | | **-210+170 =-40** |

$$\overline{X} = X + \frac{\Sigma fdx}{N}$$

$X = 155$

$\Sigma fdx = -40$

$N = 40$

$$\overline{X} = 155 + \frac{-40}{40}$$

$= 155 + (-1) = 154$

Arithmetic Mean $= 154$

**Illustration 17 :** From the following data calculate Arithmetic

MeanClass      0-10    10-20 20-30 30-40 40-50 50-60

Frequency    12    18    27    20    17    6

**Solution**

| Class | Frequency | Mid Point of X | dx | fdx |
|---|---|---|---|---|
| 0-10 | 12 | 5 | -30 | -360 |
| 10-20 | 18 | 15 | -20 | -360 |
| 20-30 | 27 | 25 | -100 | 270 |
| 30-40 | 20 | 35 | 0 | 0 |
| 40-50 | 17 | 45 | +10 | +170 |
| 50-60 | 6 | 55 | +20 | +120 |
| | **100** | | | **-990+290 =-700** |

$$X = X + \frac{\Sigma fdx}{N}$$

$$X = 35$$

$$\Sigma fdx = -700$$

$$N = 100$$

$$\overline{X} = 35 + \frac{-700}{100}$$

$$= 35 + (-7) = 28$$

Arithmeic Mean = 28

## Step Deviation Method :

$$\overline{X} = X + \frac{\Sigma fDx}{N} \, xC$$

X = assumed MeanC = common factor

$\Sigma fDx$ = sum of multiplies of Dx with

frequenciesN = total frequency

**Illustration 18 :** Calculate Arithmetic Mean.

| Class | 35-40 | 40-45 | 45-50 | 50-55 | 55-60 | 60-65 | 65-70 | 70-75 | 75-80 | 80-85 |
|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 7 | 8 | 12 | 26 | 32 | 42 | 42 | 15 | 17 | 9 |

| Class | Frequency | Mid Point of X | dx | Dx 5 | fDx |
|---|---|---|---|---|---|
| 35-40 | 7 | 37.5 | -20 | -4 | -28 |
| 40-45 | 8 | 42.5 | -15 | -3 | -24 |
| 45-50 | 12 | 47.5 | -10 | -2 | -24 |
| 50-55 | 26 | 52.5 | -5 | -1 | -26 |
| 55-60 | 32 | 57.5 | 0 | 0 | 0 |
| 60-65 | 42 | 62.5 | +5 | +1 | +42 |
| 65-70 | 42 | 67.5 | +10 | +2 | +84 |
| 70-75 | 15 | 72.5 | +15 | +3 | +45 |
| 75-80 | 17 | 77.5 | +20 | +4 | +68 |
| 80-85 | 9 | 82.5 | +25 | +5 | +45 |
| | **210** | | | | **-102+284 =+182** |

$$\overline{X} = X + \frac{\Sigma fDx}{N} \, xC$$

X = 57.5, C = 5

$\Sigma fDx = +182$

N = 210

$$\overline{X} = 57.5 + \frac{182}{210} \, x \, 5$$

$$\overline{X} = 57.5 + \frac{910}{210} = 57.5 + 4.33$$

$\overline{X} = 61.83$

## When Mid points are given :

If mid points are given take Midpoints directly and calculate arithmetic mean.

### Illustration 19 :

| Mid Points | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Frequency | 2 | 60 | 101 | 152 | 205 | 155 | 79 | 40 | 1 |

**Solution**

| Mid Points | Frequency | dx | fDx |
|---|---|---|---|
| 1 | 2 | -4 | -8 |
| 2 | 60 | -3 | -180 |
| 3 | 101 | -2 | -202 |
| 4 | 152 | -1 | -152 |
| 5 | 205 | 0 | 0 |
| 6 | 155 | +1 | +155 |
| 7 | 79 | +2 | +158 |
| 8 | 40 | +3 | +120 |
| 9 | 1 | +4 | +4 |
| | **795** | | **-105** |

$$\overline{X} = X + \frac{\Sigma fdx}{N}$$

$$X = 5$$

$$\Sigma fdx = -105$$

$$N = 795$$

$$\overline{X} = 5 + \frac{-105}{795}$$

$$\overline{X} = 5 + (-0.132) = 4.868$$

$$\overline{X} = 4.868$$

**Inclusive Method :**

When the data is given in inclusive form, then it is not necessary to adjust the classes for calculating arithmetic mean. it is because the mid value, remains the same whether the adjustment is made or not.

**Illustration 20** : Calculate Arithmetic Mean from the following.

| Marks | 1-5 | 6-10 | 11-15 | 16-20 | 21-25 | 26-30 | 31-35 | 36-40 | 41-45 | 46-50 | 51-55 | 56-60 | 61-65 |
|-------|-----|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| No. of Students | 4 | 8 | 27 | 48 | 57 | 81 | 86 | 77 | 49 | 36 | 20 | 5 | 2 |

**Solution :**

| Class | Frequency | Mid Point of X | dx | Dx $\frac{}{5}$ | fDx |
|-------|-----------|----------------|-----|-----|-----|
| 35-40 | 7 | 37.5 | -20 | -4 | -28 |
| 40-45 | 8 | 42.5 | -15 | -3 | -24 |
| 45-50 | 12 | 47.5 | -10 | -2 | -24 |
| 50-55 | 26 | <u>52.5</u> | -5 | -1 | -26 |
| 55-60 | 32 | 57.5 | 0 | 0 | 0 |
| 60-65 | 42 | 62.5 | +5 | +1 | +42 |
| 65-70 | 42 | 67.5 | +10 | +2 | +84 |
| 70-75 | 15 | 72.5 | +15 | +3 | +45 |
| 75-80 | 17 | 77.5 | +20 | +4 | +68 |
| 80-85 | 9 | 82.5 | +25 | +5 | +45 |
| | **210** | | | | **-102+284 =+182** |

| Marks | No. of Students | Mid Points | dx | Dx (5) | fdx |
|-------|-----------------|------------|------|--------|------|
| 1-5   | 4  | 3  | -30 | -6 | -24  |
| 6-10  | 8  | 8  | -25 | -5 | -40  |
| 11-15 | 27 | 13 | -20 | -4 | -108 |
| 16-20 | 48 | 18 | -15 | -3 | -144 |
| 21-25 | 57 | 23 | -10 | -2 | -104 |
| 26-30 | 81 | 28 | -5  | -1 | -81  |
| 31-35 | 86 | <u>33</u> | 0 | 0 | 0 |
| 36-40 | 77 | 38 | +5  | +1 | +77  |
| 41-45 | 49 | 43 | +10 | +2 | +98  |
| 46-50 | 36 | 48 | +15 | +3 | +108 |
| 51-55 | 20 | 53 | +20 | +4 | +80  |
| 56-60 | 5  | 58 | +25 | +5 | +25  |
| 61-65 | 2  | 63 | +30 | +6 | +12  |
|       | **500** |  |  |  | **-111** |

$$\overline{X} = X + \frac{\Sigma fdx}{N} \times C$$

$$X = 33, C = 5$$

$$\Sigma fdx = -111$$

$$N = 500$$

$$\overline{X} = 33 + \frac{-111}{500} \times 5$$

$$\overline{X} = 33 + (-1.11)$$

$$\overline{X} = -31.89$$

## Unequal Classes :

If the given classes are not equal, no need to change the class to calculate arithmetic mean.

## Illustration 21 :

From the following data calculate Arithmetic Mean.

Revenue Rs. : Below 50   50-70      70-100     100-110 110-120 120-above

No.of Persons         8     12          20      30      7     3

**Solution :**

| Revenue | No. of Persons | Mid Points | dx | fdx |
|---------|----------------|------------|-----|-----|
| 30-50 | 8 | 40 | -45 | -360 |
| 50-70 | 12 | 60 | -25 | -300 |
| 70-100 | 20 | <u>85</u> | 0 | 0 |
| 100-110 | 30 | 105 | +20 | +600 |
| 110-120 | 7 | 115 | +30 | +210 |
| 120-130 | 3 | 125 | +40 | +120 |
| | **80** | | | **-660+930 =+270** |

$$\overline{X} = X + \frac{\Sigma fdx}{N}$$

$$X = 85, \ \Sigma fdx = 270, \qquad N = 80$$

$$\overline{X} = 85 + \frac{270}{80}$$

$$\overline{X} = 85 + 3.38$$

$$\overline{X} = 88.38$$

## Open-End Classes :

Open-end classes are those in which lower limit of the first class and the upper limit of the last class are not known. In such case we cannot find out the Arithmetic Mean unless we make an assumption about the unknown limits. the assumption would naturally depend upon the class interval following the first class and preceding the last class.

**Illustration 22 :**

Class            below 50  50-100  100-150  150-200  200-250  above 250

Frequency        57       256     132      25       10        12

**Solution :**

| C lass | Frequency | Mid Points | d x | fd x |
|---|---|---|---|---|
| Below 50 | 57 | 25 | -100 | -5700 |
| 50-100 | 256 | 75 | -50 | -12800 |
| 100-150 | 132 | <u>125</u> | 0 | 0 |
| 150-200 | 25 | 175 | +50 | 1250 |
| 200-250 | 10 | 225 | +100 | 1000 |
| 250-300 | 12 | 275 | +150 | 1800 |
| | **492** | | | **-18500+4050 =-14450** |

$$\overline{X} = X + \frac{\Sigma fdx}{N}$$

$$X = 125, \qquad \Sigma fdx = -14450, N = 492$$

$$\overline{X} = 125 + \frac{-14450}{492}$$

$$\overline{X} = 125 + (-29.36) = 95.64$$

$$\overline{X} = 95.64$$

**Illustration 23 :**

From the following data calculate Arithmetic Mean.

| Income Rs. | 35-40 | 40-45 | 45-50 | 50-55 | 55-60 | 60-75 | 75-90 | 90-100 | 100-120 |
|---|---|---|---|---|---|---|---|---|---|
| No. of Persons | 6 | 7 | 13 | 15 | 16 | 14 | 11 | 9 | 9 |

**Solution :**

First Class        40

Second Class 40 - 45 Difference - 5

So take the first class difference also as 5.

Now the first class lower limit is = Upper limit - 5

$$= 40 - 5 = 35$$

| Income Rs. | No. of Persons | Mid Points | dx | fdx |
|---|---|---|---|---|
| 35-40 | 6 | 37.5 | -17.5 | -105.0 |
| 40-45 | 7 | 42.5 | -12.5 | -87.5 |
| 45-50 | 13 | 47.5 | -7.5 | -97.5 |
| 50-60 | 15 | <u>55.0</u> | 0 | 0 |
| 60-75 | 16 | 67.5 | +12.5 | 200 |

| 75-90 | 14 | 82.5 | +27.5 | 385 |
|---|---|---|---|---|
| 90-100 | 11 | 95.0 | +40.5 | 440 |
| 100-120 | 9 | 110.0 | +55.0 | 495 |
| 120-140 | 9 | 130.0 | +75.0 | 675 |
| | **100** | | | **+2195-290 =+1905** |

$$\overline{X} = X + \frac{\Sigma fdx}{N}$$

$X = 55, \Sigma fdx = 1905, N = 100$

$$X = 55 + \frac{1905}{100}$$

$\overline{X} = 55 + 19.05 = 74.05$

$\overline{X} = 74.05$

## Arithmetic Mean with Cumulative Frequency Distribution :

When the data is given in the form of 'more than' or 'less than', 'above' or 'below' for all items, in the series, it is called cumulative frequency distribution. To calculate arithmetic mean, construct class by taking difference of two given limits. Get general frequency by substracting cumulative frequency.

## Less than Cumulative Frequency :

**Illustration 24 :**

Following are the marks of 80 B.Com., Students in Statistics. Calculate their average mark.

| Marks | No. of Students |
|---|---|
| Less than 10 | 7 |
| Less than 20 | 18 |
| Less than 30 | 35 |
| Less than 40 | 56 |
| Less than 50 | 70 |
| Less than 60 | 76 |
| Less than 70 | 80 |

**Solution**

| Marks | No. of Students | General Frequency | Mid Points | dx | fdx |
|---|---|---|---|---|---|
| Less than 10 | 7 | 7 | 5 | -30 | -210 |
| Less than 20 | 18 | 11(18-7) | 15 | -20 | -220 |
| Less than 30 | 35 | 17(35-18) | 25 | -10 | -170 |
| Less than 40 | 56 | 21(56-35) | <u>35</u> | 0 | 0 |
| Less than 50 | 70 | 14(70-56) | 45 | +10 | +140 |
| Less than 60 | 76 | 6(76-70) | 55 | +20 | +120 |
| Less than 70 | 80 | 4(80-76) | 65 | +30 | +120 |
| | | **80** | | | **-600+380 =-220** |

$$\overline{X} = X + \frac{\Sigma fdx}{N}$$

$$X = 35, \ \Sigma fdx = -220, \ N = 80$$

$$\overline{X} = 35 + \frac{-200}{80}$$

$$\overline{X} = 35 + (-2.5) = 32.5$$

$$\overline{X} = 32.5$$

**Illustration 25 :** From the following data of 240 students marks calculate Arithmetic Mean.

| Marks | No. of Students |
|---|---|
| Less than 10 | 25 |
| Less than 20 | 40 |
| Less than 30 | 60 |
| Less than 40 | 75 |
| Less than 50 | 95 |
| Less than 60 | 125 |
| Less than 70 | 190 |
| Less than 80 | 240 |

**Solution**

| Marks | No. of Students | General Frequency | Mid Points | dx | Dx (C-10) | fDx |
|---|---|---|---|---|---|---|
| 0-10 | 25 | 25 | 5 | -30 | -3 | -75 |

| 10-20 | 40 | 15(40-25) | 15 | -20 | -2 | -30 |
| 20-30 | 60 | 20(60-40) | 25 | -10 | -1 | -20 |
| 30-40 | 75 | 15(75-60) | 35 | 0 | 0 | 0 |
| 40-50 | 95 | 20(95-75) | 45 | +10 | +1 | +20 |
| 50-60 | 125 | 30(125-95) | 55 | +20 | +2 | +60 |
| 60-70 | 190 | 65(190-125) | 65 | +30 | +3 | +195 |
| 70-80 | 240 | 50(240-190) | 75 | +40 | +4 | +200 |
|  |  | **240** |  |  |  | **-125+475 =+350** |

$$\overline{X} = X + \frac{\Sigma fDx}{N} \times C$$

$X = 35, \Sigma fDx = 350, N = 240, C = 10$

$$\overline{X} = 35 + \frac{350}{240} \times 10$$

$\overline{X} = 35 + 1.45 \times 10 = 35 + 14.5 = 49.58;$

$\overline{X} = 49.58$

## Illustration 26 :

Calculate Arithmetic Mean from the following data.

| Marks | No. of Students |
|---|---|
| Less than 10 | 15 |
| Less than 20 | 35 |
| Less than 30 | 60 |
| Less than 40 | 84 |
| Less than 50 | 96 |
| Less than 60 | 127 |
| Less than 70 | 198 |
| Less than 80 | 250 |

**Solution :**

| Marks | No. of Students | General Frequency | Mid Points | dx | Dx (C-10) | fDx |
|---|---|---|---|---|---|---|
| 0-10 | 15 | 15 | 5 | -30 | -3 | -45 |
| 10-20 | 35 | 20(35-15) | 15 | -20 | -2 | -40 |

| 20-30 | 60 | 25(60-35) | 25 | -10 | -1 | -25 |
| 30-40 | 84 | 24(84-60) | <u>35</u> | 0 | 0 | 0 |
| 40-50 | 96 | 12(96-84) | 45 | +10 | +1 | +12 |
| 50-60 | 127 | 31(127-96) | 55 | +20 | +2 | +62 |
| 60-70 | 198 | 71(198-127) | 65 | +30 | +3 | +213 |
| 70-80 | 250 | 52(250-198) | 75 | +40 | +4 | +208 |
|  |  | **250** |  |  |  | **-110+495 = +385** |

$$\overline{X} = X + \frac{\Sigma fDx}{N} \times C$$

$X = 35, \Sigma fDx = 385, N = 250, C = 10$

$$\overline{X} = 35 + \frac{385}{250} \times 10$$

$\overline{X} = 35 + 1.54 \times 10 = 35 + 15.4 = 50.4$

$\overline{X} = 50.4$

**More than Cumulative Frequency :Illustration 27 :**

Calculate average weight from the following data.

| Weight (Pounds) | No. of Persons |
|---|---|
| More than 100 | 400 |
| More than 110 | 300 |
| More than 120 | 170 |
| More than 130 | 100 |
| More than 140 | 80 |
| More than 150 | 50 |

**Solution**

| Marks | No. of Students | General Frequency | Mid Points | dx | Dx (C-10) | fDx |
|---|---|---|---|---|---|---|
| 100-110 | 400 | 100(400-300) | 105 | -20 | -2 | -200 |
| 110-120 | 300 | 130(300-170) | 115 | -10 | -1 | -130 |

| 120-130 | 170 | 70(170-100) | 125 | 0 | 0 | 0 |
|---------|-----|-------------|-----|-----|-----|-----|
| 130-140 | 100 | 20(100-80) | 135 | +10 | +1 | +20 |
| 140-150 | 80 | 30(80-50) | 145 | +20 | +2 | +60 |
| 150-160 | 50 | 50 | 155 | +30 | +3 | +150 |
| | | **400** | | | | **-330+230 =-100** |

$$\overline{X} = X + \frac{\Sigma fDx}{N} x C$$

X = 125, $\Sigma fDx$ = -100, N = 400, C = 10

$$\overline{X} = 125 + \frac{-100}{400} x 10$$

$\overline{X}$ = 125 + (-0.25) x 10 = 125 - 2.5 = 122.5

$\overline{X}$ = 122.5

**Illustration 28 :**

Calculate Arithmetic Mean from the following information

| Height (cm) | More than 75 | More than 85 | More than 95 | More than 105 | More than 115 | More than 125 | More than 135 | More than 145 |
|-------------|------|------|------|------|------|------|------|------|
| No. of Persons | 214 | 212 | 189 | 140 | 77 | 32 | 8 | 7 |

**Solution :**

| Heights | No. of Students | General Frequency | Mid Points | dx | Dx (C-10) | fDx |
|---------|-----------------|-------------------|------------|-----|-----------|-----|
| 75-85 | 214 | 2(214-212) | 80 | -30 | -3 | -6 |
| 85-95 | 212 | 23(212-189) | 90 | -20 | -2 | -46 |
| 95-105 | 189 | 49(189-140) | 100 | -10 | -1 | -49 |
| 105-115 | 140 | 63(140-77) | 110 | 0 | 0 | 0 |
| 115-125 | 77 | 45(77-32) | 120 | +10 | +1 | +45 |
| 125-135 | 32 | 24(32-8) | 130 | +20 | +2 | +48 |

| 135-145 | 8 | 1(8-7) | 140 | +30 | +3 | +3 |
|---|---|---|---|---|---|---|
| 145-155 | 7 | 7 | 150 | +40 | +4 | +28 |
| | | **214** | | | | **-162** |

$$\overline{X} = X + \frac{\Sigma fDx}{N} \, xC$$

$X = 110, \ \Sigma fDx = -162, \ N = 214, \ C = 10$

$$\overline{X} = 110 + \frac{-162}{214} \, x10$$

$\overline{X} = 110 + (-0.75) \times 10 = 110 - 7.5 = 102.5$

$\overline{X} = 102.5$

**Illustration 29** :

Calculate Arithmetic Mean.

| Class | More than 0 | More than 10 | More than 20 | More than 30 | More than 40 |
|---|---|---|---|---|---|
| Frequency | 40 | 36 | 28 | 15 | 5 |

**Solution :**

| Class | Frequency | General Frequency | Mid Points | dx | Dx (C-10) | fDx |
|---|---|---|---|---|---|---|
| 0-10 | 40 | 4(40-36) | 5 | -20 | -2 | -8 |
| 10-20 | 36 | 8(36-28) | 15 | -10 | -1 | -8 |
| 20-30 | 28 | 13(28-15) | 25 | 0 | 0 | 0 |
| 30-40 | 15 | 10(15-5) | 35 | +10 | +1 | +10 |
| 40-50 | 5 | 5 | 45 | +20 | +2 | +10 |
| | | **40** | | | | **-16+20 =+4** |

$$\overline{X} = X + \frac{\Sigma fDx}{N} \, xC$$

$X = 25, \ \Sigma fDx = +4, \ N = 40, \ C = 10$

$$\overline{X} = 25 + \frac{4}{40} \, x10$$

$\overline{X} = 25 + 0.1 \times 10 = 25 + 1 = 26$

$\overline{X} = 26$

## Correcting Incorrect Values :

It sometimes happens that due to an oversight or mistake in copying certain wrong items are taken while calculating mean. The problem is how to find out the correct mean. The process is from incorrect $\Sigma$ x deduct wrong items and add correct items and then divide the correct $\Sigma$ x by the number of observations. The result is correct mean.

### Illustration 30 :

It mean of 200 items was 50. Later on it was discovered that two items were misread as 92 and 8 instead of 192 and 88. Find the correct mean.

$\overline{X} = \dfrac{\Sigma X}{N}$

$\Sigma X = N \times \overline{X}$

$N = 200$

$\overline{X} = 50$

$\Sigma X = 200 \times 50$

Incorrect $\Sigma X = 1000$

Correct $\Sigma X$ = Incorrect $\Sigma X$ - Wrong items + Correct Items

Correct $\Sigma X = 10000 - (92 + 8) + (192 + 88)$

$\qquad = 9900 + 280 = 10180$

Correct $\Sigma X = 10180$

Correct Mean $= \dfrac{10180}{200} = 50.9$

$\overline{X} = 50.9$

### Illustration 31 :

Following are the results of 50 students who appeared for an

examination.Marks    4    5    6    7    8    9

No. of Students Passed    8    10    9    6    4    3

Average of 50 students marks are 5.16. Find out average marks of students who failed.

### Solution :

| Marks(X) | No. of Students Passed (f) | xf |
|----------|----------------------------|----|
| 4 | 8 | 32 |

| 5 | 10 | 50 |
|---|----|----|
| 6 | 9 | 54 |
| 7 | 6 | 42 |
| 8 | 4 | 32 |
| 9 | 3 | 27 |
|   |   | 237 |

Total marks of 50 students  = 5.16 x 50 = 258.00

Total marks of 40 students  = 237

Total marks of 10 students who failed = 258 - 237 = 21

Arithmetic Mean = $\dfrac{21}{10}$ = 2.1

## Missing Figures :
### Illustration 32 :

From the following information find out missed value, where the average salary is Rs. 115.86.

| Salary (X) | 110 | 112 | 113 | 117 | x | 125 | 128 | 130 |
|------------|-----|-----|-----|-----|---|-----|-----|-----|
| No. of Persons(f) | 25 | 17 | 13 | 15 | 14 | 8 | 6 | 2 |

**Solution**

| Salary (Rs.) | No. of Persons | xf |
|--------------|----------------|-----|
| 110 | 25 | 2750 |
| 112 | 17 | 1904 |
| 113 | 13 | 1469 |
| 117 | 15 | 1755 |
| x | 14 | 14x |
| 125 | 8 | 1000 |
| 128 | 6 | 768 |
| 130 | 2 | 260 |
| | **100** | **9906 + 14x** |

$\overline{X} = \dfrac{\Sigma xf}{N}$

N = 100

$\Sigma xf = 9906 + 14x$

Arithmetic Mean = 115.86

$$115.86 = \frac{9906 + 14x}{100}$$

$$14x = 11586 - 9906 \qquad 1680$$

$$x = \frac{}{14}$$

$$x = 120$$

$$= 120$$

## Illustration 33 :

From the following data find out missed frequency. if the average income is Rs. 19.92.

| Revenue (Rs.) | 4-8 | 8-12 | 12-16 | 16-20 | 20-24 | 24-28 | 28-32 | 32-36 | 36-40 |
|---|---|---|---|---|---|---|---|---|---|
| No. of Persons | 11 | 13 | 16 | 14 | x | 9 | 17 | 6 | 4 |

**Solution :**

| Revenue (Rs.) | No. of Persons | Mid Points | xf |
|---|---|---|---|
| 4-8 | 11 | 6 | 66 |
| 8-12 | 13 | 10 | 130 |
| 12-16 | 16 | 14 | 224 |
| 16-20 | 14 | 18 | 252 |
| 20-24 | x | 22 | 22x |
| 24-28 | 9 | 26 | 234 |
| 28-32 | 17 | 30 | 510 |
| 32-36 | 6 | 34 | 204 |
| 36-40 | 4 | 38 | 152 |
| | **90 + x** | | **1772 + 22x** |

Arithmetic Mean $\overline{X} = 19.92$ N $= 90 + x$

$$\Sigma xf = 1772 + 22x$$

$$\overline{X} = \frac{\Sigma xf}{N}$$

$$19.92 = \frac{1772 + 22x}{90 + x}$$

$$1772 + 22x = (19 + x)\ 19.92$$

$$1772 + 22x = 1792.80 + 19.92x$$

$$22x - 19.92x = 1792.80 - 1772$$

$$2.08x = 2080$$

$$x = \frac{2080}{2.08}$$

$$x = 10$$

## 5.6 COMBINED AVERAGE :

If Arithmetic Mean and the number of items of two or more than two related groups are given, the combined average of these groups can be calculated by applying the following formula.

$$\overline{X}_{1\ 2\ 3\ ....\ n} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2 + N_3\overline{X}_3 .................... + N_n X_n}{N_1 + N_2 + N_3 ........ + N_n}$$

$\overline{X}_{1\ 2\ 3}..............n$ = Combined mean of the groups

$\overline{X}_1$ = Arithmetic Mean of first group

$\overline{X}_2$ = Arithmetic Mean of Second group

$\overline{X}_3$ = Arithmetic Mean of third group

$\overline{X}_n$ = Arithmetic Mean of nth group

$N_1$ = Number of items in the first group

$N_2$ = Number of items in the second group

$N_3$ = Number of items in the third group

$N_n$ = Number of items in the nth group

**Illustration 34 :**

The Mean height of 25 male workers in a factory is 61 cm, and the mean height of 35 femaleworkers in the same factory is 58 cm. Find the combined mean height of 60 workers in the factory.

**Solution** :

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

Where

$\overline{X}_1 = 61$

$\overline{X}_2 = 58$

$N_1 = 25$

$$N_2 = 35$$

$$\overline{X}_{12} = \frac{(25 \times 61) + (35 \times 58)}{25 + 35}$$

$$\overline{X}_{12} = \frac{1525 + 2030}{60}$$

$$\overline{X}_{12} = \frac{3555}{60}$$

$$\overline{X}_{12} = 59.25$$

**Illustration 35 :**

The mean of wages in factory A of 100 workers is Rs. 720 per week. The mean wages of 30 female workers in the factory was Rs.650 per week. Find out average wage of male workers in the factory.

**Solution :**

$$N = 100, \ N_1 = 30, \ X_1 = 650$$

$$X_2 = ? \ N_1 + N_2 = N_{30} + N_2 =$$

$$100, \ N_2 = 70$$

$$\overline{X}_{12} = 720$$

$$720 = \frac{30 \times 650 + 70 \, X_2}{100}$$

$$72000 = 19500 + 70 \, X_2$$

$$70 \, X_2 = 52500$$

$$X_2 = \frac{52500}{70}$$

$$X_2 = 750$$

## 5.7 WEIGHTED ARITHMETIC MEAN :

One of the limitations of the arithmetic mean is that it gives equal importance to all the items. but there are cases where the relative importance of the different items is not the same. When this is so, we compute weighted arithmetic mean. The term weight stands for the relative importance of the different items. The formula for computing weighted arithmetic mean is:

$$\overline{X}_w = \frac{\Sigma wx}{\Sigma w}$$

$\overline{X}_w$ = Weighted arithmetic mean

W = Weights

X = values

**Illustration 36:**

From the following data calculate weighted Arithmetic Mean.

| Variables | 80 | 75 | 67 | 86 | 35 |
|-----------|----|----|----|----|----|
| Weights   | 2  | 3  | 4  | 5  | 6  |

**Solution :**

| S.No | X | W | XW |
|------|-----|-----|------|
| 1 | 80 | 2 | 160 |
| 2 | 75 | 3 | 225 |
| 3 | 67 | 4 | 268 |
| 4 | 86 | 5 | 430 |
| 5 | 35 | 6 | 210 |
| | **343** | | **1293** |

$$\overline{X}_w = \frac{\Sigma wx}{\Sigma w}$$

$$\overline{X}_w = \frac{1293}{20}$$

$$\overline{X}_w = 64.65$$

**Illustration 37 :**

From the following results of Three Universities calculate weighted Arithmetic Mean.

| Examination | A No. of students | | B No. of Students | | C No. of Students | |
|-------------|----------|-------------|----------|-------------|----------|-------------|
| | % passed | In hundreds | % Passed | in hundreds | % Passed | in hundreds |
| M.A. | 70 | 5 | 75 | 4 | 75 | 6 |
| M.Sc. | 85 | 4 | 80 | 3 | 65 | 4 |
| M.Com. | 80 | 6 | 65 | 5 | 70 | 5 |
| B.A. | 75 | 7 | 85 | 6 | 80 | 7 |
| B.Sc. | 65 | 5 | 75 | 4 | 85 | 5 |
| B.Com. | 75 | 8 | 70 | 5 | 75 | 5 |

**Solution :**

| Examinati-on | University A No. of students | | | University B No. of Students | | | C No. of Students | | |
|--------------|---------------|------------------------|-----|---------------|------------------------|-----|---------------|------------------------|-----|
| | % passed (X) | In hundreds (W) | XW | % passed (X) | in hundreds (W) | XW | % passed (X) | in hundreds (W) | XW |
| M.A. | 70 | 5 | 350 | 75 | 4 | 700 | 75 | 6 | 450 |
| M.Sc. | 85 | 4 | 340 | 80 | 3 | 240 | 65 | 4 | 260 |
| M.Com. | 80 | 6 | 480 | 65 | 5 | 325 | 70 | 5 | 350 |
| B.A. | 75 | 7 | 525 | 85 | 6 | 510 | 80 | 7 | 560 |
| B.Sc. | 65 | 5 | 325 | 75 | 4 | 300 | 85 | 5 | 425 |
| B.Com. | 75 | 8 | 600 | 70 | 5 | 350 | 75 | 5 | 375 |
| | **450** | **35** | **2620** | **450** | **27** | **2025** | **450** | **32** | **2420** |

Weighted Arithmetic Mean $\bar{X}w = \dfrac{\Sigma wx}{\Sigma w}$

University A $= \dfrac{2620}{35} = 74.86\%$

University B $= \dfrac{2025}{27} = 75.00\%$

University C $= \dfrac{2420}{32} = 75.64\%$

**Illustration 38 :**

From the results of the College X and Y. State which of them is better and why.

| Name of the Exam | College X | | | College Y | |
|---|---|---|---|---|---|
| | Appeared | Passed | | Appeared | Passed |
| M.A. | 300 | 250 | 1000 | 800 | |
| M.Com. | 500 | 450 | 1200 | 950 | |
| B.A. | 2000 | 1500 | 1000 | 700 | |
| B.Com. | 1200 | 750 | 800 | 500 | |

**Solution :**

| Name of the Exam | Appeared | College X % of Pass | $W_1X_1$ | Appeared | College Y % of Pass | $W_1X_1$ |
|---|---|---|---|---|---|---|
| M.A. | 300 | $\dfrac{250}{300} \times 100 = 83.3$ | 24,990 | 1000 | $\dfrac{800}{1000} \times 100 = 80.0$ | 80,000 |
| M.Com. | 500 | $\dfrac{450}{500} \times 100 = 90.0$ | 45,000 | 1200 | $\dfrac{950}{1200} \times 100 = 79.17$ | 95,004 |
| B.A. | 2000 | $\dfrac{1500}{2000} \times 100 = 75.0$ | 150000 | 1000 | $\dfrac{700}{1000} \times 100 = 70.0$ | 70,000 |
| B.Com. | 1200 | $\dfrac{750}{1200} \times 100 = 62.5$ | 75000 | 800 | $\dfrac{500}{800} \times 100 = 62.5$ | 50,000 |
| | 4000 | | 2,94,990 | 4000 | | 2,95,004 |

Weighted Arithmetic Mean for College X =        $\frac{\Sigma w_1 x_1}{\Sigma w_1}$

Weighted Arithmetic Mean for College Y =        $\frac{\Sigma w_2 x_2}{\Sigma w_2}$

$= \dfrac{294990}{4000}$

$= \dfrac{295004}{4000}$

$= 73.747$

$= 73.751$

Average pass percentage of College Y > College X.So, College Y is better.

## Merits and Limitations of Arithmetic Mean :

**Merits :** Arithmetic Mean is most commonly used average in practice because of its advantages. Following are important Merits of Arithmetic Mean.

1. It is very simple to understand and calculate.

2. Arithmetic Mean is affected by the value of each and every item in the series.

3. Arithmetic Mean is defined by a rigid mathematical formula.

4. Arithmetic Mean is useful for algebraic treatment.  it is better than median Mode, Geometric Mean and Harmonic Mean.

5. Arithmetic Mean is relatively stable.  It does not fluctuate much when repeated samplesare taken from one and the same population.

6. Arithmetic Mean is a calculated value and is not based on position in the series.

## Demerits

1. The value of Arithmetic Mean depends on each and every item of the series.  The valueof average is affected by the extreme items, either very small or very large.

2. In open-end classes, the value of mean cannot be calculated without making assumption regarding the size of the class interval of the open - end classes.

3. In case the distribution is U shaped, then mean is not likely to serve a useful purpose.So it is not a good measure always.

## 5.9  SUMMARY :

Thus, it is most widely used measure for representing the entire data. To a layman, it is average but for a statistician, it is called 'arithmetic mean'.  It is calculated by adding values of all the items and dividing their total by the number of items.  In case of discrete and continuous series, the values of the frequencies are taken into account. Following figure depicts the Calculation of arithmetic Mean.

**Arithmetic Mean ($\overline{X}$)**

```
                    Individual Series        Discrete Series        Continuous Series

              Direct Method  Shortcut    Direct Method  Shortcut    Direct Method
                    Shortcut Method                     Method          Method
```
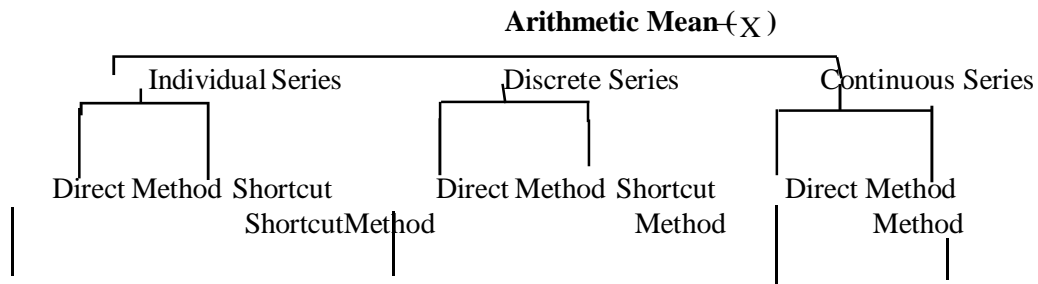
## 5.10 EXERCISE :

1. What is an average? What are its objectives.

2. Explain requisites of a good average.

3. Define Arithmetic Mean and Explain its merits and demerits.

4. From the following data calculate Arithmetic Mean.

   Monthly Income Rs. : 200, 300, 330, 400, 500, 600, 400, 700, 740, 560, 440

   (Ans. : 470)

5. Find out Arithmetic mean.

   | Wages | 3 | 5 | 8 | 10 | 12 | 15 |
   |---|---|---|---|---|---|---|
   | No. of Workers | 4 | 10 | 12 | 8 | 4 | 2 |

   (Ans. : 7.90)

6. Calculate Arithmetic mean.

   | Class | 15-25 | 25-35 | 35-45 | 45-55 | 55-65 | 65-75 |
   |---|---|---|---|---|---|---|
   | Frequencies | 20 | 30 | 40 | 50 | 60 | 70 |

   (Ans. : 40.2)

7. Findout Arithmetic Mean.

   | Class | 2-3 | 4-5 | 6-7 | 8-9 | 10-11 | 12-13 |
   |---|---|---|---|---|---|---|
   | Frequency | 20 | 43 | 50 | 30 | 18 | 10 |

   (Ans. : 6.66)

8. Calculate Arithmetic Mean.

   | Class | Frequency |
   |---|---|
   | Below 5 | 5 |
   | 10 | 9 |
   | 15 | 17 |
   | 20 | 29 |
   | 25 | 45 |
   | 30 | 60 |
   | 35 | 70 |
   | 40 | 78 |
   | 45 | 83 |
   | 50 | 85 |

   (Ans. 24.25)

9.Calculate Arithmetic Mean

| Income | No. of Persons |
|---|---|
| More than 10 | 72 |
| More than 20 | 67 |
| More than 30 | 59 |
| More than 40 | 50 |
| More than 50 | 36 |
| More than 60 | 21 |
| More than 70 | 9 |
| More than 80 | 3 |

(Ans. 49.03)

10. Calculate Arithmetic Mean.

| Wages | No. of Workers |
|---|---|
| 5 | 7 |
| 10 | 8 |
| 15 | 12 |
| 20 | 13 |
| 25 | 18 |
| 30 | 14 |
| 35 | 11 |
| 40 | 8 |
| 45 | 5 |
| 50 | 4 |

(Ans. : 2.25)

11. Calculate the number of students against the class 30-40 of the following data where $\bar{X}$ 28.

| Marks | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 |
|---|---|---|---|---|---|---|
| No. of Students | 12 | 18 | 27 | 2 | 17 | 6 |

12. Average weight of 150 students in a class is 60 kgs. Average weight of Boys of that class is 70 kgs, and girls average weight is 55 kgs.  Find out number of boys and girls ofthat class.  (Ans. : 50, 100)

**- Dr. K. Kanaka Durga**

# LESSON - 6
# AVERAGES - II MEDIAN

**OBJECTIVES :**

After studying this lesson you should be able to understand -

1. What is Median.

2. What are it merits and limitations.

3. How to compute Median.

**STRUCTURE OF THE LESSON :**

## 6.1   INTRODUCTION :

The median is one of the measures of central value. One of the most important objects of statistical analysis is to get one single value that describes the characteristic of the entire mass of un widely data such value is called the central value or an average. As distinct from the Arithmetic mean which is calculated from the value of every item in the series, the median is that is called a positional average. The term 'position' refers to the place of a value in a series. The place of the median in a series is such that an equal number of items lie on either side of it. For example, if the income of five persons is 2,800, 2820, 2880, 2885, 2890, then the median income would be Rs. 2,880. median is thus the central value of the distribution or the value that divides the distribution into two equal parts.

## 6.2   MEANING AND DEFINITION :

The median by definition is the middle value of the distribution. Whenever the median is given as a measure, one half of the items in the distribution have a value.

Thus the median divides the distribution into two equal parts. If there are even number of items in a series there is no actual value exactly in the middle of the series and as such the median is indeterminate. In such a case the median is arbitrarily taken to be halfway between the two middle values.

## 6.3    CALCULATION OF MEDIAN :

Median is calculated in three series such as Individual series, Discrete series and Continuous series.

Median

Individual Series          Discrete Series          Continuous Series
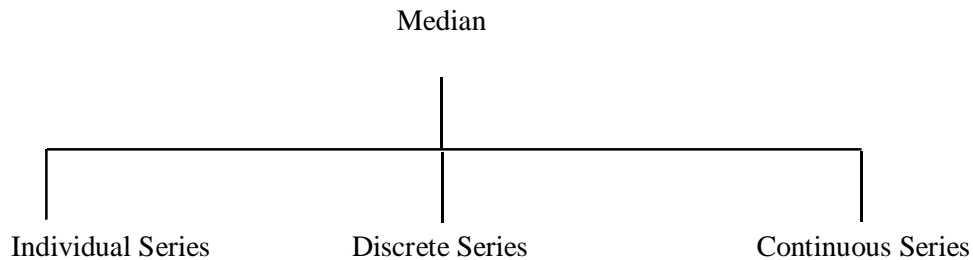
### 6.3.1 Individual Series :

Following is the procedure to calculate Median in Individual Series.

1. Arrange the data in ascending or descending order of magnitude

2. Apply the formula

$$M = \text{size of } \frac{N+1}{2} \text{ th item.}$$

N = No. of items

**Illustration 1:**

Obtain the value of median from the following data.

| Values | 391 | 384 | 407 | 672 | 522 | 777 | 753 | 2488 | 1490 |
|--------|-----|-----|-----|-----|-----|-----|-----|------|------|

**Solution :**

| Values | 384 | 391 | 407 | 522 | <u>672</u> | 753 | 777 | 1490 | 2488 |
|--------|-----|-----|-----|-----|-----|-----|-----|------|------|

$$\text{Median} = \text{size of } \frac{N+1}{2} \text{ th item}$$

N = 9 (Number of items)

$$\text{Size of } \frac{9+1}{2} \text{ th item}$$

$$= \frac{10}{2} \text{ th item}$$

5th item i.e. Median = 672

**Illustration 2 :** From the wages of 11 workers, Calculate Median.

| Values | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|--------|---|---|---|---|---|---|---|---|---|----|----|
| Wages | 60 | 55 | 45 | 70 | 75 | 80 | 50 | 90 | 95 | 100 | 85 |

**Solution**

| Values | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|--------|---|---|---|---|---|---|---|---|---|----|----|
| Wages | 45 | 50 | 55 | 60 | 70 | _75_ | 80 | 85 | 90 | 95 | 100 |

$$\text{Median} = \text{size of } \frac{N+1}{2} \text{ th item}$$

N = 11 (Number of items)

$$\text{Size of } \frac{11+1}{2} \text{ th item}$$

$$\frac{12}{2} \text{ th item}$$

6th item i.e. Median = 75

**Illustration 3 :**

| Marks | 30 | 27 | 26 | 35 | 37 | 40 | 25 | 45 | 47 |
|-------|----|----|----|----|----|----|----|----|----|

**Solution**

| Marks in Ascending Order | 25 | 26 | 27 | 30 | _35_ | 37 | 40 | 45 | 47 |
|--------------------------|----|----|----|----|------|----|----|----|----|

$$\text{Median} = \text{size of } \frac{N+1}{2} \text{ th item}$$

N = 9 (Number of items)

$$\text{Size of } \frac{9+1}{2} \text{ th item}$$

$$= \frac{\frac{10}{2}}{} \text{ th item}$$

= 5th item i.e. Median = 35

If the number of items was odd and therefore, it is not possible to determine the middle value. When the number of items is even for example, if in the above case the number of items is 8

then median would be the value of $\frac{8+1}{2}$ = 4.5 th item for finding out the value of 4.5th item we shall

take the average of 4th and 5th items i.e. $\text{Median} = \frac{\text{4th item+ 5th item}}{2}$

**Illustration 4 :**

Calculate median income from the following data.

Income Rs. :  891 884    991    907    1072    922    1277    1153    2488    1490

Solution

Income

$$\text{Median = size of } \frac{N+1}{2} \text{ th item}$$

Rs.

N = 10 (Number of items)

884

891

$$\text{Size of } \frac{10+1}{2} \text{ th item}$$

907

222 th item       $\frac{11}{2}$

991

1072              5th item + 6th item Median

$$= \frac{}{2}$$

1153

2488           Hence Median Income is Rs.1031.5

**Illustration 5 :**

Calculate Median mark from the following data.

Marks : 40      45      31      75      81      57      63      92      35      21

**Solution :**

| Marks | Marks in Ascending Order |
|-------|------|
| 40 | 21 |
| 45 | 31 |
| 31 | 35 |
| 75 | 40 |
| 81 | 45 |
| 57 | 57 |
| 63 | 63 |
| 92 | 75 |
| 35 | 81 |
| 21 | 92 |

$$\text{Median} = \text{size of } \frac{N+1}{2} \text{ th item}$$

N = 10 (Number of items)

$$\text{Size of } \frac{10+1}{2} \text{ th item}$$

th item

$$= \frac{\text{5th item} + \text{6th item}}{2} \text{ Median}$$

$$= \frac{45 + 57}{2} = \frac{102}{2} \quad 57 = 51$$

Median = 51

**Illustration 6 :**

From the following wages. Calculate Median.

| Wages (Rs.) | 60 | 55 | 45 | 70 | 75 | 80 | 50 | 90 | 95 | 100 |
|-------------|----|----|----|----|----|----|----|----|----|-----|
| | | | | | 95 | | | | | |

Solution

| Wages (Rs.) | Wage s in Ascending Order (Rs.) |
|-------------|------|
| | 60    45 |
| 55 | 50 |
| 45 | 55 |
| 70 | 60 |
| 75 | 70 |
| 80 | 85 |
| 50 | 80 |

$$\text{Median} = \text{size of } \frac{N+1}{2} \text{ th item}$$

N = 12 (Number of items)

Size of $\frac{12+1}{2}$ th item

75 th item 6.5th item

| 90 | 85 | Median = $\dfrac{\text{6th item + 7th item}}{2}$ |

| 95 | 90 |

| 100 | 95 | $\dfrac{75 + 80}{2}$ $\dfrac{155}{2}$ = 77.5 |

| 95 | 100 |

| 110 | 110 |

## 6.3.2 Discrete Series :

$$\text{Median = Size of } \dfrac{N+1}{2} \text{ th item}$$

N = Total of the frequency.

Steps of calculate Median.

1. Arrange data in ascending order or descending order.

2. Find out Cumulative Frequency

3. Apply the formula.

**Illustration 7 :**

From the following data find out the Value of Median :

| Income (Rs.) | No. of Persons |
|---|---|
| 1600 | 24 |
| 1650 | 26 |
| 1580 | 16 |
| 1700 | 20 |
| 1750 | 6 |
| 1680 | 30 |

**Solution**

| Income in Ascending Order (Rs.) | No. of Persons | Cumulative Frequency |
|---|---|---|
| 1580 | 16 | 16 |

| 1600 | 24 | 40 |
| 1650 | 26 | 66 |
| 1680 | 30 | 96 |
| 1700 | 20 | 116 |
| 1750 | 6 | 122 |
| | 122 | |

$$\text{Median} = \text{size of } \frac{N+1}{2} \text{ th item}$$

$$N = 122$$

$$M = \text{Size of } \frac{122+1}{2} \text{ th item}$$

$$\frac{123}{2} \text{ th item} = 61.5 \text{ th item.}$$

Size of 61.5th item = 1650. Hence Median 1650.

**Illustration 8 :**

From the following heights of 100 students calculate Median Height.

| Height Cm. : | 155 | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 |
|---|---|---|---|---|---|---|---|---|---|---|
| No. of Students | 3 | 7 | 9 | 12 | 13 | 17 | 16 | 14 | 7 | 2 |

**Solution :**

| Height | No. of Students | Cumulative cm. Frequency |
|---|---|---|
| 155 | 3 | 3 |
| 156 | 7 | 10 |
| 157 | 9 | 19 |
| 158 | 12 | 31 |
| 159 | 13 | 44 |
| 160 | 17 | 61 |
| 161 | 16 | 77 |

| 162 | 14 | 91 |
|-----|----|----|
| 163 | 7 | 98 |
| 164 | 2 | 100 |
| | 100 | |

$$\text{Median} = \text{size of } \frac{N+1}{2} \text{ th item}$$

$$N = 100$$

$$M = \text{Size of } \frac{100+1}{2} \text{ th item}$$

$$\frac{101}{2} = \text{th item} = 50.5 \text{ th item.}$$

Size of 50.5th item = 61 of Cumulative Frequency.

Median = Corresponding value of 61 is 160 Median

= 160

Illustration :

From the following Weights. Calculate Median Weight.

| Weight (P) | 70 | 100 | 180 | 150 | 80 | 120 | 200 | 250 | 170 | 90 |
|------------|----|-----|-----|-----|----|-----|-----|-----|-----|----|
| No. of Persons | 40 | | 20 | 45 | 25 | 38 | 35 | 50 | 22 | 15 | 30 |

**Solution :**

| Weight | No. of Persons | Captive Frequency |
|--------|----------------|-------------------|
| 70 | 20 | 20 |
| 80 | 35 | 55 |
| 90 | 40 | 95 |
| 100 | 45 | 140 |
| 120 | 50 | 190 |

| | | |
|---|---|---|
| 150 | 38 | 228 |
| 170 | 30 | 258 |
| 180 | 25 | 283 |
| 200 | 22 305 | 15 |
| 250 | 320 | |
| | 320 | |

Median = size of $\dfrac{N+1}{2}$ th item

N = 320

Size of $\dfrac{320+1}{2}$ th item

$\dfrac{321}{2}$ th item = 160.5 th item.

Size of 160.5th item lies in 190.

Corresponding Value of 190 = 120.

Median = 120

### 6.3.3 Continuous Series :

In continuous series calculation of Median follows the following steps.

1. Determine he particular class in which the value of median lies with the help of $m = \dfrac{N}{2}$.

2. Apply the Principle

$$M = l_1 + \dfrac{l_2 - l_1}{f_1} \times m - c$$

$l_1$ = Lower limit of the Median Class

$l_2$ = Upper limit of the Median Class $f_1$ =

Frequency of the Median Class

$m$ = Value of $\dfrac{N}{2}$ nd item.

$c$ = Cumulative Frequency of the class preceding the Median Class.

### Illustration 10 :

From the following data calculate Median.

| Values | 0-20 | 20-40 | 40-60 | 60-80 | 80-100 | 100-120 | 120-140 | 140-160 |
|--------|------|-------|-------|-------|--------|---------|---------|---------|
| Value Frequency | 1 | 14 | 35 | 85 | 90 | 45 | 18 | 2 |

**Solution**

| Values | Value Frequency | Cumulative Frequency |
|--------|-----------------|----------------------|
| 0-20 | 1 | 1 |
| 20-40 | 14 | 15 |
| 40-60 | 35 | 50 |
| 60-80 | 85 | 135 |
| 80-100 | 90 | 225 |
| 100-120 | 45 | 270 |
| 120-140 | 18 | 288 |
| 140-160 | 2 | 290 |

$$\underline{290}$$

$$M = l_1 + \frac{l_2 - l_1}{f_1} \times m - c$$

$$M = 80 + \frac{35}{90} \times 10$$

$$M = 80 + \frac{206}{96}$$

$$M = 80 + 2.2 = 82.2.$$

M    = 82.2

**Illustration 11 :**

Calculate Median from the following marks.

| Marks | No. of Students |
|-------|-----------------|
| 30-32 | 2 |
| 32-34 | 9 |

| | |
|---|---|
| 34-36 | 25 |
| 36-38 | 30 |
| 38-40 | 45 |
| 40-42 | 62 |
| 42-44 | 39 |
| 44-46 | 20 |
| 46-48 | 11 |
| 48-50 | 3 |

**Solution :**

| Marks | No. of Students | Cumulative Frequency |
|---|---|---|
| 30-32 | 2 | 2 |
| 32-34 | 9 | 11 |
| 34-36 | 25 | 36 |
| 36-38 | 30 | 66 |
| 38-40 | 45 | 115 |
| 40-42 | 62 | 177 |
| 42-44 | 39 | 216 |
| 44-46 | 20 | 236 |
| 46-48 | 11 | 247 |
| 48-50 | 3 | 250 |
| | **250** | |

$m = \dfrac{250}{2}$

$125 \quad M = l_1 + \dfrac{250}{2} \dfrac{l_2 - l_1}{} \times m - c$

$f_1 \quad l_1 = 40 \quad l_2 = 42 \quad f_1 = 62 \quad m = 125$

c = 115

$$M = 40 + x \frac{42 - 40}{2} \frac{125 - 115}{62}$$

$$M = 40 + \frac{6220}{62} \times 10$$

$$M = 40 +$$

$$M = 40 + 0.3 = 40.3$$

$$M = 40.3$$

**Illustration 12 :**

Calculate Median from the following data.

| Class | Frequency | Cumulative Frequency |
|-------|-----------|----------------------|
| 15-N25 | 4 | 4 |
| 25-35 | 11 | 15 |
| 35-45 | 19 | 34 |
| 45-55 | 14 | 48 |
| 55-65 | 0 | 48 |
| 65-75 | 2 | 50 |
| | **50** | |

m = 2

N = 50

$$m = \frac{50}{2} = 25$$

$$l^2 - l_1$$

$$M = l_1 + \underline{\quad} \times m - c$$
$$\phantom{M = l_1 +} f_1$$

$l_1 = 35$

$l_2 = 45$

$f_1 = 19$

$m = 25$

$c = 15$

$$M = 35 + \cfrac{45 - 35}{19}$$
$$\times 25 - 15$$

$$M = 35 + \cfrac{10}{19} \times 10$$

$$M = 35 + \cfrac{100}{19}$$

$$M = 35 + 5.2 = 40.2$$

$N = 40.2$

## Illustration 13 :

From the following data calculate Median Profit.

| Profit(Rs.) | No. of Traders |
|---|---|
| 1999.5-2999.5 | 20 |
| 2999.5-3999.5 | 45 |
| 3999.5-4999.5 | 70 |
| 4999.5-5999.5 | 50 |
| 5999.5-6999.5 | 28 |
| 6999.5-7999.5 | 22 |
| 7999.5-8999.5 | 15 |

**Solution :**

| Profit(Rs.) | No. of Traders | Cumulative Frequency |
|---|---|---|
| 1999.5-2999.5 | 20 | 20 |
| 2999.5-3999.5 | 45 | 65 |
| 3999.5-4999.5 | 70 | 135 |
| 4999.5-5999.5 | 50 | 185 |
| 5999.5-6999.5 | 28 | 213 |
| 6999.5-7999.5 | 22 | 235 |
| 7999.5-8999.5 | 15 | 250 |
| | **250** | |

N = 250

$$m \quad \frac{250}{2} = = 125$$

$$M = l_1 + \underline{\quad} l_2 - l_1 \, x \, m - c$$
$$f_1$$

$$M = 3999.5 + \frac{4999.5 - 3999.5}{70} \, x \, 125\text{-}65$$

$$M = 3999.5 + \frac{1000}{70} \, x \, 60$$

$$M \quad = \quad \frac{6000}{7}$$
3999.5 +

M = 3999.5 + 857.1 = 4856.6

M = 4856.6

**6.3.3.1 Inclusive (Class) Series :**

When the classes are in inclusive series, change the classes into exclusive form. To change into exclusive form take the difference between upper limit of first class and the lower limit of next class. Divide the difference by two. Subtract the difference from lower limits and add to the upper limits.

**Example** :      11-20

                21-30

                31-40

Difference between 21-20 is 1. It is divided by two i.e. $\dfrac{1}{2} = 0.5$.

20 + 0.5 = 20.5 Upper limit

21 - 0.5 = 20.5 lower limit.

**Illustration 14 :**

From the following data calculate Median.

| Class | Frequency |
|-------|-----------|
| 11-20 | 21 |
| 21-30 | 19 |
| 31-40 | 60 |
| 41-50 | 42 |
| 51-60 | 24 |
| 61-70 | 18 |
| 71-80 | 15 |

**Solution :**

| Class | Frequency | c.f |
|-------|-----------|-----|
| 11-20 | 21 | 21 |
| 21-30 | 19 | 40 |
| 31-40 | 60 | 100 |
| 41-50 | 42 | 142 |
| 51-60 | 24 | 166 |

| 61-70 | 18 | 184 |
| 71-80 | 15 | 199 |

$$m = \frac{N}{2} \text{ nd item}$$

$$m = \frac{199}{2} = 99.5$$

$$M = l_1 + \frac{l_2 - l_1}{} \times m - c$$

$$f_1 \quad l_1 = 30.5 \quad l_2 = 40.5 \quad f_1 = 60$$

$$m = 99.5 \quad c = 40$$

$$\frac{40.5 - 30.5}{70}$$

$$M = 30.5 + 10 \qquad \times 99.5 - 40$$

$$M = 30.5 + \frac{}{60} \times 59.5$$

$$M = 30.5 + \frac{595}{60}$$

$$M \qquad = 30.5 + 9.9 = 40.4$$

**Illustration 15 :**

From the following Incomes of 9,990 persons. Calculate Median Income.

| Revenue (Rs.) | No. of Persons |
|---|---|
| 0-9 | 2756 |
| 10-19 | 2124 |
| 20-29 | 1677 |
| 30-39 | 1481 |
| 40-49 | 1021 |
| 50-59 | 610 |
| 60-69 | 245 |
| 70-79 | 67 |
| 80-89 | 6 |

| 90-99 | 3 |
|-------|---|
|       | **9990** |

**Solution :**

| Revenue (Rs.) | No. of Persons | Exclusive Class | Cumulative Frequency |
|---------------|----------------|-----------------|----------------------|
| 0-9   | 2756 | -0.5-9.5  | 2756 |
| 10-19 | 2124 | 9.5-19.5  | 4880 |
| 20-29 | 1677 | 19.5-29.5 | 6557 |
| 30-39 | 1481 | 29.5-39.5 | 8039 |
| 40-49 | 1021 | 39.5-49.5 | 9059 |
| 50-59 | 610  | 49.5-59.5 | 9669 |
| 60-69 | 245  | 59.5-69.5 | 9914 |
| 70-79 | 67   | 69.5-79.5 | 9981 |
| 80-89 | 6    | 79.5-89.5 | 9987 |
| 90-99 | 3    | 89.5-99.5 | 9990 |
|       | **9990** |       |      |

$$m = \frac{}{2}$$

$$N = 9990$$

$$m = \frac{9990}{2} = 4995$$

$$M = l_1 + \frac{l_2 - l_1}{f_1} \times m - c$$

$$l_1 = 19.5$$

$$l_2 = 29.5$$

$f_1 = 1677$   $m = 4995$

$c = 4880$

$$M = 19.5 + \frac{29.5 - 19.5}{1677} \text{ x } 4995 - 4880$$

$$M = 19.5 \quad \frac{10}{1677} \quad + \text{ x } 15$$

$$M = 19.5 + \frac{1150}{1677}$$

$M = 19.5 + 0.6 = 20.1$

$M = 20.1$

### 6.3.3.2 Un equal Classes :

When the class intervals are unequal the frequencies need not be adjusted to make the class intervals
equal and the same formula for interpolation can be applied.

**Illustration 16:**

| Marks | No. of Students |
|-------|-----------------|
| 0-10  | 5               |
| 10-30 | 15              |
| 30-60 | 30              |
| 60-80 | 8               |
| 80-90 | 2               |

**Solution :**

| Marks | No. of Students | Cumulative Frequency |
|-------|-----------------|----------------------|
| 0-10  | 5               | 5                    |

| | | |
|---|---|---|
| 10-30 | 15 | 20 |
| 30-60 | 30 | 50 |
| 60-80 | 8 | 58 |
| 80-90 | 2 | 60 |

$$\text{Median} = \text{Size of } \frac{N}{2} \text{nd item 2 N}$$

$$= \text{nd item} = 30\text{th item.}$$

Median lies in the Class 30-60.

$$\text{Median} = l_1 + \frac{l_2 - l_1}{f_1} \times m - c$$

$$l_1 = 30 \quad l_2 = 60 \quad f_1 = 30$$

$$M = 30 + \frac{60 - 30}{30} \times 30 - 20$$

$$M = 30 + \frac{30}{30} \times 10$$

$$M = 30 + \frac{300}{30}$$

$$M = 30 + 10 = 40$$

$$M = 40$$

### 6.3.3.3 When Mid Points are given :

When Mid points are given in the problem construct class by taking difference between two mid points.

Example :

| Mid Point | Class | | |
|---|---|---|---|
| 115 | 110-120 | (115-5) | (115+5) |
| 125 | 120-130 | (125-5) | (125+5) |
| 135 | 130-140 | (135-5) | (135+5) |

125 - 115 = 10,

10 / 2 = 5,

115 - 5 = 110,

115 + 5= 120

**Illustration 17 :**

Compute Median from the following data.

| Mid Value | Frequency |
|-----------|-----------|
| 115 | 6 |
| 125 | 25 |
| 135 | 48 |
| 145 | 72 |
| 155 | 116 |
| 165 | 60 |
| 175 | 38 |
| 185 | 22 |
| 195 | 3 |

**Solution :**

| Mid Value | Frequency | Cumulative Frequency |
|-----------|-----------|----------------------|
| 115 | 6 | 6 |
| 125 | 25 | 31 |
| 135 | 48 | 79 |
| 145 | 72 | 151 |
| 155 | 116 | 267 |
| 165 | 60 | 327 |
| 175 | 38 | 365 |
| 185 | 22 | 387 |
| 195 | 3 | 390 |
| | **390** | |

Difference between two mid points is 115 - 125 = 10 Divide

10 by 2 = i.e. 5

Mid Value -5 = Lower Limit = 115 - 5 = 110

Mid Value +5 = Upper limit = 115 + 5 = 120

Class is = 110 - 120

$$\text{Median} = \text{Size of } \frac{N}{2}\text{nd item}$$

$$= \frac{380}{2} \text{ th item} = 195 \text{ th item}$$

Median class = 150 - 160

$$\text{Median} = l_1 + \frac{l_2 - l_1}{f_1} \times m - c$$

$$M = 150 + \frac{160 - 150}{116} \times 195 - 151$$

$$M = 150 + \frac{10}{116} \times 44$$

$$M = 150 + \frac{440}{116}$$

$$M = 150 + 3.79; \qquad M = 153.79$$

**Illustration 18 :**

| Mid Value | Frequency |
|---|---|
| 1 | 3 |
| 2 | 60 |
| 3 | 101 |
| 4 | 152 |
| 5 | 205 |
| 6 | 155 |
| 7 | 79 |
| 8 | 40 |

**Solution :**

Difference between two mid points is 1

Divide 1 by 2 = 0.5

Mid value - 0.5 = Lower Limit

Mid value + 0.5 = Upper Limit 1

- 0.5 = 0.5 - Lower Limit

1 + 0.5 = 1.5 Upper Limit

| Class | Frequency | Cumulative Frequency |
|---|---|---|
| 0.5 - 1.5 | 3 | 3 |
| 1.5 - 2.5 | 60 | 63 |
| 2.5 - 3.5 | 101 | 164 |
| 3.5 - 4.5 | 152 | 316 |
| 4.5 - 5.5 | 205 | 521 |
| 5.5 - 6.5 | 155 | 676 |
| 6.5 - 7.5 | 79 | 755 |
| 7.5 - 8.5 | 40 | 795 |
| | **795** | |

$$\text{Median} = \text{Size of } \frac{N}{2} \text{nd item}$$

$$= \frac{795}{2} \text{ th item} = 397.5 \text{ th item.}$$

$$\text{Median} = l_1 + \frac{\frac{2}{} - l_1}{f_1} \times m - c$$

$$M = 4.5 + \frac{5.5 - 4.5}{205} \times 397.5 - 316$$

$$M = 4.5 + \frac{1}{205} \times 81.5$$

$$M = 4.5 + \frac{81.5}{205}$$

M = 4.5 + 0.4 ; M = 4.9

### 6.3.3.4 Cumulative Frequency - Median (Less than, More than Methods) :

When the data are given in the form of 'Less than', 'More than'. The given frequency is cumulative frequency. It is necessary to convert it into simple frequency distribution.

**Illustration 19 :**

From the following 655 Students. Calculate Median.

| Values | Frequency |
|---|---|
| Less than 5 | 29 |
| Less than 10 | 224 |
| Less than 15 | 465 |
| Less than 20 | 582 |
| Less than 25 | 634 |
| Less than 30 | 644 |
| Less than 35 | 650 |
| Less than 40 | 653 |
| Less than 45 | 655 |

**Solution :**

| Values | Frequency | Cumulative Frequency | |
|---|---|---|---|
| 0 - 5 | 29 | 29 | 29 |
| 5 - 10 | 195 | (224-29) | 224 |
| 10 - 15 | 241 | (465-224) | 465 |
| 15 - 20 | 117 | (582-468) | 582 |

| | | | |
|---|---|---|---|
| 20 - 25 | 52 | (634-582) | 634 |
| 25 - 30 | 10 | (655-634) | 644 |
| 30 - 35 | 6 | (650-644) | 650 |
| 35 - 40 | 3 | (653-650) | 653 |
| 40 - 45 | 2 | 655-653) | 655 |
| | **655** | | |

Median = Size of $\dfrac{N}{2}$ nd item

N = 655 655

m = ___ = 327.5
     2

Median = $l_1 + \dfrac{l_2 - l_1}{f_1}$ x m - c

$l_1$ = 10 $l_2$ = 15

$f_1$ = 241 m = 327.5 c = 224

M = 10 $\dfrac{15 - 10}{241}$ + x 327.5 - 224

M = 10 $\dfrac{5}{241}$ + x 103.5

M = $\dfrac{517.5}{241}$
10 +

M= 10 + 2.14. M = 12.14

**Illustration 20 :** From the following Marks. Calculate Median Mark.

| No. of Students | 100 | 90 | 80 | 60 | 32 | 20 | 13 | 5 |
|---|---|---|---|---|---|---|---|---|
| Marks | Less than 80 | Less than 70 | Less than 60 | Less than 50 | Less than 40 | Less than 30 | Less than 20 | Less than 10 |

**Solution :**

| Marks | No. of Students | Marks | No. of Students | Cumulative Frequency |
|---|---|---|---|---|
| Less than 10 | 5 | 0-10 | 5 | 5 |
| Less than 20 | 13 | 10-20 | 13 | 8 |
| Less than 30 | 20 | 20-30 | 20 | 7 |
| Less than 40 | 32 | 30-40 | 32 | 12 |
| Less than 50 | 60 | 40-50 | 60 | 28 |
| Less than 60 | 80 | 50-60 | 80 | 20 |
| Less than 70 | 90 | 60-70 | 90 | 10 |
| Less than 80 | 100 | 70-80 | 100 | 10 |
| | | | | **100** |

Median = Size of $\dfrac{N}{2}$ nd item

N = 100

$m = \dfrac{100}{2} = $ 50th item

Median = $l_1 + \dfrac{l_2 - l_1}{f_1} \times m - c$

$l_1 = 40, l_2 = 50, f_1 = 28, m = 50, c = 32$

$M = 40 + \dfrac{50 - 40}{28} \times 50 - 32$

$M = 40 + \dfrac{10}{28} \times 18$

$M = 40 + \dfrac{180}{28}$

$M = 40 + 6.4$

$M = 46.42$

**More Than :**

**Illustration 21 :**

From the following data. Calculate Median.

| Class | Frequency |
|---|---|
| More than 90 | 51 |
| More than 100 | 49 |
| More than 110 | 49 |
| More than 120 | 43 |
| More than 130 | 37 |
| More than 140 | 17 |
| More than 150 | 5 |

**Solution :**

| Class | Frequency | Cumulative Frequency |
|---|---|---|
| 90-100 | 2 | 2 |
| 100-110 | 0 | 2 |
| 110-120 | 6 | 8 |
| 120-130 | 6 | 14 |
| 130-140 | 20 | 34 |
| 140-150 | 12 | 46 |
| 150-160 | 5 | 51 |
| | **51** | |

$$\text{Median} = \text{Size of } \frac{N}{2} \text{nd item}$$

N = 51

$$m = \frac{51}{2} \text{ nd item}$$

i.e. 25.5th item

$$\text{Median} = l_1 + \frac{2 - l_1}{} \times \text{m - c} \frac{l}{f_1}$$

$l_1 = 130$

$$M = 130 + \frac{140 - 130}{20} \times 25.5 - 14$$

$$M = 130 + \frac{10}{20} \times 11.5$$

$$M = 130 + \frac{115}{20}$$

$$M = 130 + 5.75$$

$$M = 135.75$$

## 6.4  CALCULATION OF MEDIAN BY GRAPHIC METHOD :

Median can be calculated by Graphic Method. This is possible with the help of ogive curves which are also known as cumulative frequency curves. Cumulative frequency curves are two types.

**a)      Less than Curve**: In order to draw these curves we have first of all to convert the ordinary frequencies into a cumulative frequency series. The frequency of all the preceding class intervals are summed up to the frequency of a class. We start with the upper limits of the classes and go on adding the frequencies. In this case of ogive curve has a rising trend.

**b)      More than Curve** : In this case the frequencies of all the succeeding classes are added to the frequency of a class. We start with the lower limits of the classes and from the cumulative frequen cies, we subtract the frequency of each class. In this case the ogive curve has a down ward trend.

**Illustration 22 :**
          From the following data. Locate Median through Graph.

| Marks | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|-------|------|-------|-------|-------|-------|-------|-------|
| No. of Students | 8 | 13 | 17 | 18 | 16 | 9 | 7 |

     **Solution**

| Marks | No. of Students | Cumulative Frequency |
|-------|-----------------|----------------------|
| 0-10 | 8 | 8 |
| 10-20 | 13 | 21 |
| 20-30 | 17 | 38 |
| 30-40 | 18 | 56 |
| 40-50 | 16 | 72 |
| 50-60 | 9 | 81 |
| 60-70 | 7 | 88 |

Cumulative Frequency Curve

Scale :

on X axis = 1 cm = 10on

Y axis = 1 cm = 10
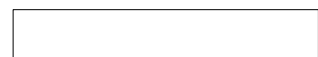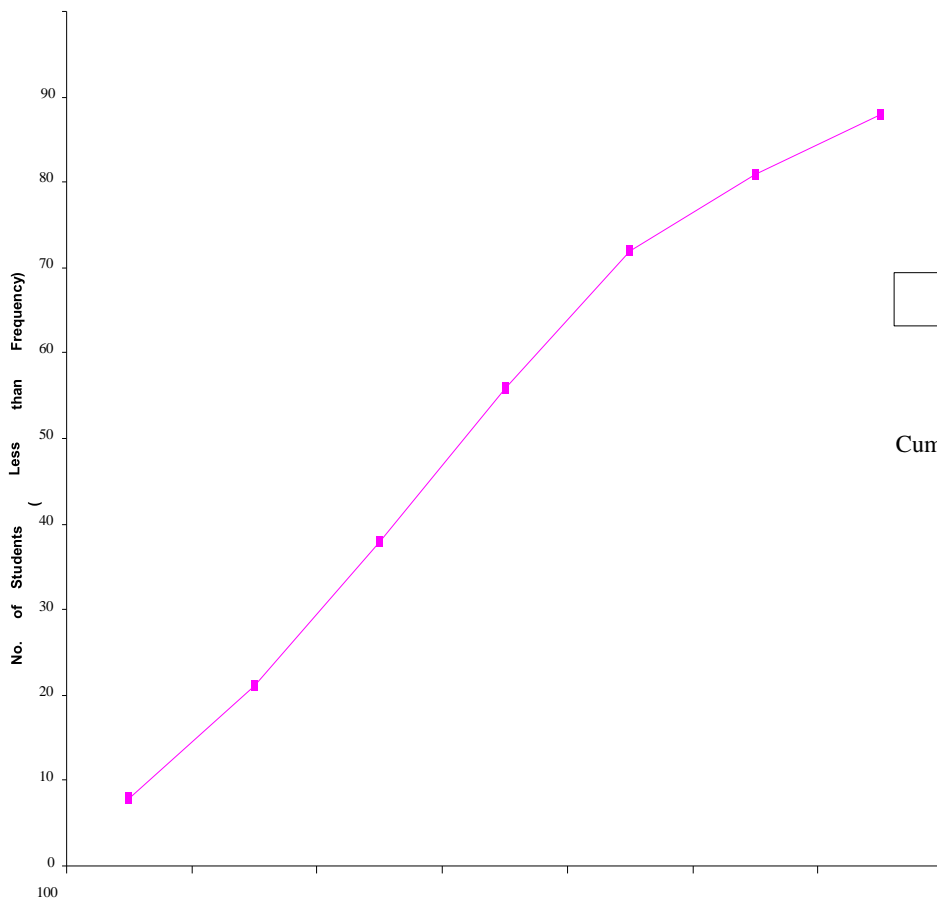


Cumulative Frequency

0-10      10-20      20-30      30-40      40-50      50-60      60-70

**Marks (Class)**

Note : Show less than frequency on 'Y' axis, and Marks on 'X' axis.

**Illustration 23 :**

From the following data show Median through Graph.

| Class | 0-5 | 5-10 | 10-15 | 15-20 | 20-25 | 25-30 | 30-35 |
|---|---|---|---|---|---|---|---|
| Frequency | 7 | 12 | 15 | 16 | 13 | 12 | 5 |

**Solution :**

| Marks | No. of Students | Cumulative Frequency |
|---|---|---|
| 0-5 | 7 | 7 |
| 5-10 | 12 | 19 |
| 10-15 | 15 | 34 |
| 15-20 | 16 | 50 |
| 20-25 | 13 | 63 |
| 25-30 | 12 | 75 |
| 30-35 | 5 | 80 |
| | **80** | |

Scale on X axis 1cm = 10, Y axis 1 cm = 10
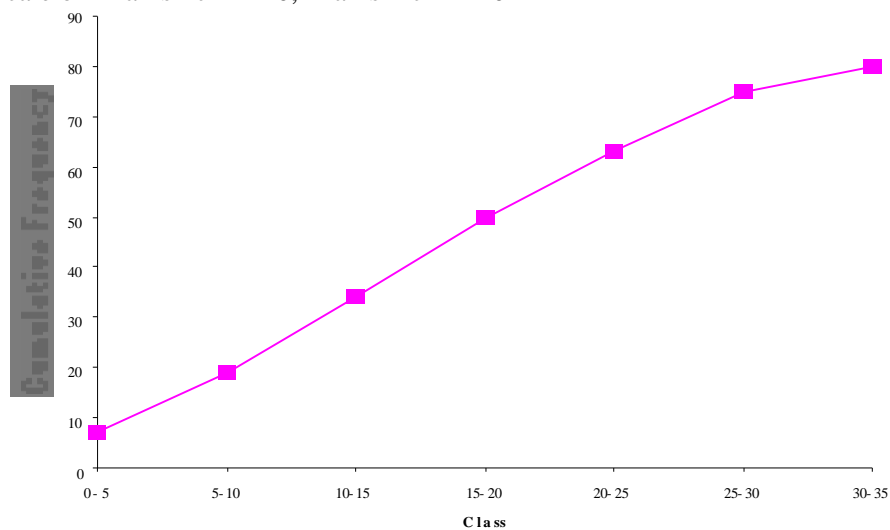
**Less than, More than Ogive Curves :**
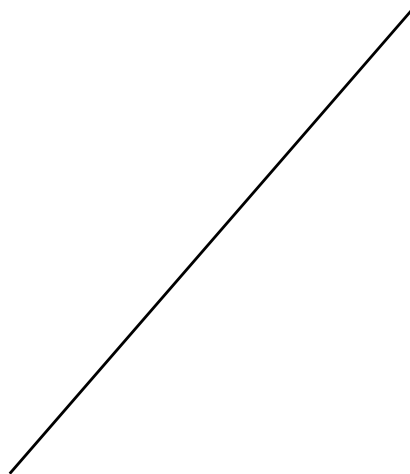


**Illustration 24 :** From the following data show the Median with the help of two Ogive Curves.

| Class | 0-100 | 100-200 | 200-300 | 300-400 | 400-500 | 500-600 |
|-------|-------|---------|---------|---------|---------|---------|
| Frequency | 13 | 15 | 17 | 18 | 16 | 21 |

**Solution :**

| X | f | c.f. Less than | c.f. More than |
|---|---|---------------|----------------|
| 0-100 | 13 | 13 | 100 |
| 100-200 | 15 | 28 | 87 |
| 200-300 | 17 | 45 | 72 |
| 300-400 | 18 | 63 | 55 |
| 400-500 | 16 | 79 | 37 |
| 500-600 | 21 | 100 | 21 |
| | **N=100** | | |
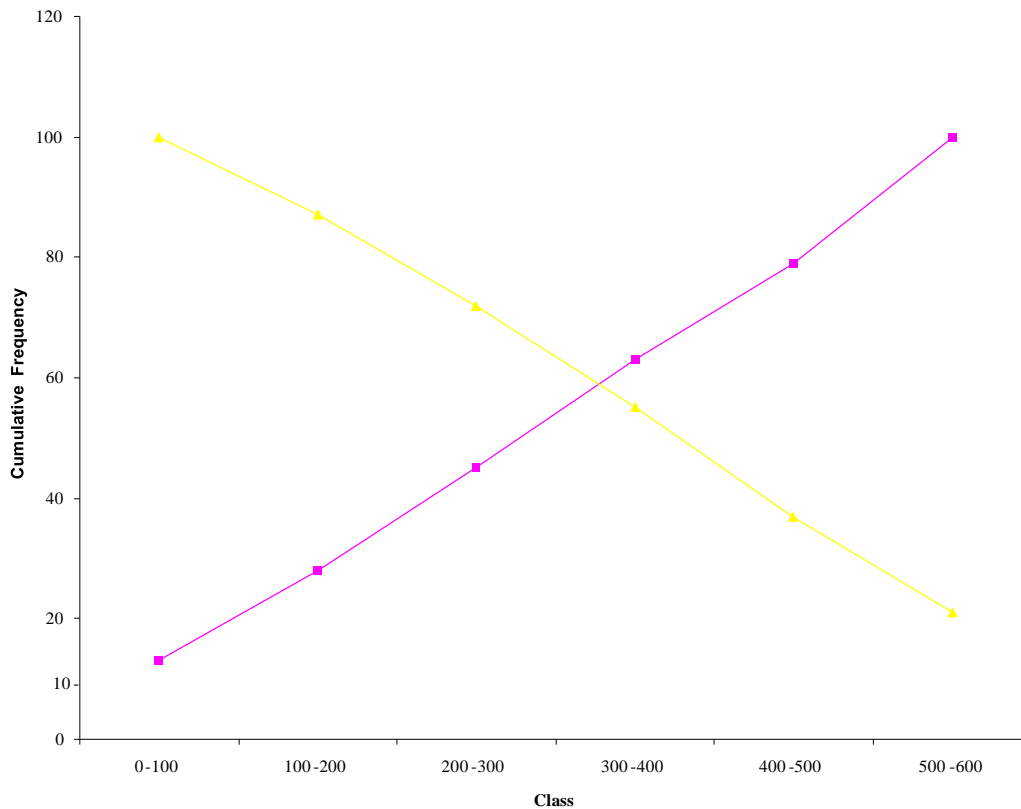
Here median is a point where two ogive curves interesect. M = 327.78

## 6.5  MERITS OF MEDIAN :

Following are the important merits or advantages of Median.

   1  It is especially useful in case of open end classes since only the position and not the values of items must be known.

   2  It is not influenced by the magnitude of extreme deviation from it.

   3  In a markedly skewed distribution such as income distribution or price distribution where the arithmetic mean would be distorted by extreme values the median is especially useful.

   4  It is most appropriate average in dealing with qualitative data.

   5  The value of median can be determined graphically whereas the value of mean cannot be graphically as certained.

## 6.6     LIMITATIONS OF MEDIAN :

        Following are the limitations of Median.

   1  For calculating median it is necessary to arrange th data. Other averages do not need any arrangements.

   2  Since it is a positional average, its value is not determined by each and every observa- tion.

   3  It is not capable of algebraic treatment.

   4  The value of median is affected more by sampling fluctuations.

5 When the number of items included in a series of data is even, the median is determined approximately as the mid-point of the two middle numbers.

## 6.7    SUMMARY :

Thus Median is the value which divides the data into two parts. it is called a positional average. The term position refers to the place of a value in a series. If there are even number of items in a series there is no actual value eractly in the middle of the series and as such the median is indeterminate Median also can be derived with the help of graph.

## 6.8    EXERCISE :

1. What is Median, Explain its merits and limitations.

2. From the following particulars calculate Median.

| Marks :75 | 24 | 42 | 57 | 63 | 49 | 91 | 12 | 8 | 20 | 35 |
|-----------|----|----|----|----|----|----|----|---|----|----|

3. From the following data find out Median.

| No. of Children | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------------|---|----|----|----|----|----|----|----|----|----|----|
| No. of Families | 7 | 12 | 75 | 89 | 80 | 47 | 35 | 23 | 12 | 13 | 5 |

4. Calculate Median from the following data?

| Wages (Rs.) | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 40 |
|-------------|----|----|----|----|----|----|----|----|----|
| No. of Workers | | 3 | 7 | 8 | 13 | 16 | 15 | 14 | 5 | 2 |

5. From the following data calculate Median.

| Marks | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 | 80-90 |
|-------|------|-------|-------|-------|-------|-------|-------|-------|-------|
| No. of Students | | 8 | 9 | 13 | 16 | 17 | 15 | 12 | 7 | 3 |

6. From the following data calculate Median.

| Class | 0-5 | 5-10 | 10-15 | 15-20 | 20-25 | 25-30 | 30-35 | 35-40 |
|-------|-----|------|-------|-------|-------|-------|-------|-------|
| Frequency | 7 | 12 | 15 | 19 | 18 | 17 | 16 | 13 |

7. Calculate Median from the following data.

| Mid Values | 12.5 | 13.0 | 13.5 | 14.0 | 14.5 | 15.0 | 15.5 | 16.0 | 16.5 | 17.0 |
|------------|------|------|------|------|------|------|------|------|------|------|
| Frequency | 13 | 19 | 23 | 27 | 28 | 31 | 26 | 21 | 18 | 17 |

8. From the following data calculate Median.

| Marks | <10 | <20 | <30 | <40 | <50 | <60 | <70 | <80 | <90 | <100 |
|---|---|---|---|---|---|---|---|---|---|---|
| No. of Students | 3 | 9 | 18 | 30 | 43 | 60 | 76 | 90 | 98 | 100 |

9. Calculate Median from the following data.

| Values | >20 | >30 | >40 | >50 | >60 | >70 | |
|---|---|---|---|---|---|---|---|
| Frequency | 65 | 63 | 40 | 40 | 18 | 7 | Ans.: 53.4 |

10. Calculate Median from the following data.

| Class | >30.0 | >32.5 | >35.0 | >37.5 | >40.0 | >42.5 | >45.0 | >47.5 | >50.0 | >52.5 | >55.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 940 | 903 | 825 | 646 | 655 | 271 | 186 | 103 | 38 | 6 | 1 |

(Ans. : 39.80)

11. From the following data find out Median.

| Revenue (Rs.) | 0-9 | 10-19 | 20-29 | 30-39 | 40-49 | 50-59 | 60-69 | 70-79 | 80-89 | 90-99 |
|---|---|---|---|---|---|---|---|---|---|---|
| No. of Persons | 2756 | 2124 | 1677 | 1481 | 1021 | 610 | 245 | 67 | 6 | 3 |

**Dr. K. Kanaka Durga**

**OBJECTIVES :**

After studying this lesson you should be able to understand.

1. What are Positional Measures.

2. How to calculate Positional Measures.

**STRUCTURE OF LESSON :**

**7.1 INTRODUCTION :**

Besides median, there are other measures which divide a series into equal parts. Important amongst these are quartiles, deciles, and percentiles.

**7.2  QUARTLES :**

Quartiles are those values of the variate which divide the total frequency into four equal parts. There are three Quartiles denoted by Q.

They are

1. Lower Quartile        -        $Q_1$

2. Upper Quartile        -        $Q_3$

3. Middle Quartile(Median) -    $Q_2$

The procedure of computing quartiles is the same as the median.

## 7.2.1 Individual Series : Quartiles

While computing Quartiles in Individual Series we add 1 to N. First

Quartile : $Q_1$ (Lower Quartile)

$$Q_1 = \text{Size of } \frac{N\ 1}{4} \text{ th item}$$

$$Q_3 = \text{Size of } \frac{N\ 1}{4} \text{ x 3rd item}$$

N = No .of items.

## Illustration

From the following data calculate First Quartile and Third Quartile.

| Wages Rs. | 45 | 50 | 60 | 55 | 75 | 70 | 85 | 90 | 90 | 95 | 100 |
|-----------|----|----|----|----|----|----|----|----|----|----|-----|

**Solution :**

Arrange data in order.

Wages 45    50    55    60    70    75    85    90    90    95    100

$$Q_1 = \frac{N+1}{4} \text{ th item}$$

N = 11 (No. of items)

$$Q_1 = \frac{11+1}{4} \text{ th item}$$

$$Q_1 = \frac{12}{4} \text{ th item}$$

$$Q_1 = \text{3rd item}$$

$$Q_1 = 55$$

Third Quartile or Upper Quartile : $Q_3$

$$Q_3 = \frac{N+1}{4} \text{ x 3rd item}$$

N = No. of

items N = 11

$$Q_3 = \frac{11+1}{4} \text{ x 3rd item}$$

$$Q_3 = \frac{12}{4} \text{ x 3rd item}$$

$$Q_3 = 3 \text{ x 3rd Item}$$

$$Q_3 = \text{9th Item}$$

$$Q_3 = 90$$

**Illustration 2 :**

From the following data. Compute $Q_1$ and $Q_3$.

| S.No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|-------|---|---|---|---|---|---|---|---|---|----|----|
| Wages (Rs.) | 61 | 64 | 66 | 67 | 68 | 69 | 70 | 73 | 74 | 75 | 76 |

**Solution**

$$Q_1 = \frac{N+1}{4} \text{ th item}$$

N = 11 (No. of items)

$$Q_1 = \frac{11+1}{4} \text{ th item}$$

$$Q_1 = \frac{12}{4} \text{ th item}$$

$$Q_1 = \text{3rd item}$$

$$Q_1 = 66$$

Third Quartile or Upper Quartile : $Q_3$

$$Q_3 = \frac{N+1}{4} \text{ x 3rd item}$$
N = No. of items N = 11

$$Q_3 = \frac{11+1}{4} \text{ x 3rd item}$$

$$Q_3 = \frac{12}{4} \text{ x 3rd item}$$

$$Q_3 = 3 \text{ x 3rd Item}$$

$$Q_3 = \text{9th Item}$$

$$Q3 = 74$$

**Illustration 3**

| S.No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Wages (Rs.) | 61 | 64 | 66 | 67 | 68 | 69 | 70 | 73 | 74 | 75 | 76 | 78 |

**Solution**

$$Q_1 = \frac{N+1}{4} \text{ th item}$$

N = 12 (No. of items)

$$Q_1 = \frac{12+1}{4} \text{ th item}$$

$$Q_1 = \frac{13}{4} \text{ th item}$$

$$Q_1 = 3.25 \text{ th item}$$

i.e. 3rd item + 0.25 x 4th item - 3rd item

66 + 0.25 x 67 - 66

66 + 0.25 x 1

66 + 0.25 = 66.25

$\square Q_1 = 66.25$

Third Quartile or Upper Quartile : $Q_3$

$$Q_3 = \frac{N+1}{4} \text{ 3rd item}$$

N = No. of

items N = 12

$$Q_3 = \frac{12+1}{4} \text{ x 3rd item}$$

$$Q_3 = \frac{13}{4} \text{ x 3rd item}$$

Q₃ = 3.25 x 3rd Item

Q₃ = 9.75 th Item

Q3 = 9th item + 0.75 x 10th item - 9th item

$\qquad$ = 74 + 0.75 x 75 - 74

$\qquad$ = 74 + 0.75 x 1

$Q_3$ = 74.75

## 7.2.2 Discrete Series :

$$Q_1 = \frac{N\ 1}{4} \text{ th item}$$

$$Q_3 = \frac{N\ 1}{4} \text{ x 3rd item}$$

N  =  Total  of  the  frequency

**Illustration**

From the following data calculate Q₁ and Q₃.

| Values | 2 | 3 | 4 | 5 | 7 | 9 | 11 | 12 |
|--------|---|---|---|---|---|---|----|----|
| Frequency | 1 | 9 | 4 | 7 | 4 | 5 | 1 | 8 |

**Solution**

| Values | Frequency | Cf |
|--------|-----------|----|
| 2 | 1 | 1 |
| 3 | 9 | 10 |
| 4 | 4 | 14 |
| 5 | 7 | 21 |
| 7 | 4 | 25 |
| 9 | 5 | 30 |
| 11 | 1 | 31 |
| 12 | 8 | 39 |

$$Q_1 = \frac{N+1}{4}$$

th item N = 39

$$Q_1 = \frac{39+1}{4} \text{ th item}$$

$$Q_1 = \frac{40}{4} \text{ th item}$$

$Q_1 = 10$ th item

$Q_1 = 3$

Third Quartile or Upper Quartile : $Q_3$

$$Q_3 = \frac{N+1}{4} \text{ x 3rd item}$$

$N = 39$

$$Q_3 = \frac{39+1}{4} \text{ x 3rd item}$$

$$Q_3 = \frac{40}{4} \text{ x 3rd item}$$

$Q_3 = 10$ x 3rd Item

$Q_3 = 30$ th Item

$\square Q_3 = 9$

**Illustration 5 :**

Find out Quartiles.

| X | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| f | 13 | 54 | 75 | 90 | 64 | 21 | 15 |

**Solution**

| X | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| f | 13 | 54 | 75 | 90 | 64 | 21 | 15 |
| Cf | 13 | 67 | 142 | 232 | 296 | 317 | 332 |

$$Q_1 = \frac{N+1}{4} \text{ th item}$$

N = 332

$$Q_1 = \frac{332+1}{4} \text{ th item}$$

$$Q_1 = \frac{333}{4} \text{ th item}$$

$Q_1 = 83.25$ th item

$Q_1 = 2$

Third Quartile or Upper Quartile : $Q_3$

$$Q_3 = \frac{N+1}{4} \text{ x 3rd item}$$

N = 332

$$Q_3 = \frac{332+1}{4} \text{ x 3rd item}$$

$$Q_3 = \frac{333}{4} \text{ x 3rd item}$$

$Q_3 = 83.25$ x 3rd Item

$Q_3 = 249.75$ th Item

$Q_3 = 4$

### 7.2.3 Continous Series

$$N \ q_1 = \frac{}{} 4 \text{ th item}$$

$$Q_1 = l_1 + \frac{l_2 - l_1}{f_1} \text{ x } q_1 - C$$

$$Q_3 = \frac{N}{4} \text{ x 3rd item}$$

$$Q_3 = l_1 + \frac{l_2 - l_1}{f_1} \text{ x } q_3 - C$$

$l_1$ = lower limit of Quantile class

$l_2$ = Upper limit of Quantile class

$f_1$ = frequency of Quantile class

$q_1$ = Value of $q_1$

C = Cumulative frequency of proceeding class of Quantile class.

## Illustration 6

From the following data calculate Quantiles.

| Age | 20-25 | 25-30 | 30-35 | 35-40 | 40-45 | 45-50 | 50-55 | 55-60 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|
| No. of Workers | 50 | 70 | 100 | 180 | 150 | 120 | 70 | 60 |

## Solution

| Age | 20-25 | 25-30 | 30-35 | 35-40 | 40-45 | 45-50 | 50-55 | 55-60 | |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|---|
| No. of Workers | 50 | 70 | 100 | 180 | 150 | 120 | 70 | 60 | 800 |
| Cumulative Frequency | 50 | 120 | 220 | 400 | 550 | 670 | 740 | 800 | |

$q_1 = \dfrac{N}{4}$

th item

$N = 800 \quad q_1 = \dfrac{800}{4}$ th item i.e. 200th item

200 th item is in cumulative frequency of 220.

The corresponding class is 30 - 35.

$Q_1 = l_1 + \dfrac{l_2 - l_1}{f_1} \times q_1 - C$

$l_1 = 30, \; l_2 = 35, \; f_1 = 100, \; q_1 = 200, \; C = 120$

$Q_1 = 30 + \dfrac{35 - 30}{100} \times 200 - 120$

$Q_1 = 30 + \dfrac{5}{100} \times 80$

$= 30 + \dfrac{400}{100} \quad Q_1 = 30 + 4 = 34 \quad q_3 = \dfrac{800}{4} \times$ 3rd item

= 200 x 3rd item = 600 th item

$$Q_3 = l_1 + \frac{l_2 - l_1}{f_1} \times q_3 - C,$$

Quartile Class = 45 - 50 $l_1 = 45$, $l_2 = 50$, $f_1 = 120$, $q_3 = 600$, C = 550

$$Q_3 = 45 + \frac{50 - 45}{120} \times 600 - 550$$

$$= 45 + \frac{5}{120} \times 50 = 45 + \frac{250}{120} = 45 + 2.08 = 47.08$$

$Q_3 = 47.08$

## Illustration 7

From the following information Calculate Quartiles.

| Values | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 |
|--------|------|-------|-------|-------|-------|-------|-------|-------|
| Frequency | 5 | 25 | 40 | 70 | 90 | 40 | 20 | 10 |

**Solution**

| Values | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 | |
|--------|------|-------|-------|-------|-------|-------|-------|-------|-----|
| Frequency | 5 | 25 | 40 | 70 | 90 | 40 | 20 | 10 | 300 |
| Cumulative Frequency | 5 | 30 | 70 | 140 | 230 | 270 | 290 | 300 | |

$$q_1 = \frac{N}{4}$$

th item N = 300 $q_1 = \frac{300}{4}$ th item i.e. 75 th item

$$Q_1 = l_1 + \frac{l_2 - l_1}{f_1} \times q_1 - C$$

$l_1 = 30$, $l_2 = 40$, $f_1 = 70$, $q_1 = 75$, C = 70

$$Q_1 = 30 + \frac{40 - 30}{70} \times 75 - 70$$

$$Q_1 = 30 + \frac{10}{70} \times 5$$

$$Q_1 = 30 + 0.71 = 30.71$$

$$q_3 = \frac{300}{4} \times \text{3rd item}$$

$$= 75 \times \text{3rd item} = 225 \text{ th item}$$

$$Q_3 = l_1 + \frac{l_2 - l_1}{f_1} \times q_3 - C, \quad \text{Quartile Class} = 45 - 50$$

$$Q_3 = 40 + \frac{50 - 40}{90} \times 225 - 145$$

$$= 40 + \frac{10}{90} \times 85 = \frac{850}{90} = 9.44$$

$$Q_3 = 49.44$$

## 7.3 DECILES :

Deciles divide the series into 10 equal parts. For any series, there are 9 deciles, as there are three quartiles for any series. Deciles range from $D_1$ to $D_9$.

## 7.3.1 Individual Series :

$$D = \frac{N + 1}{10} \times \text{Required decile}$$

N = Number of Items.

## Illustration 8 :

From the following data calculate 8th decile.

| Marks | 11 | 12 | 14 | 18 | 22 | 26 | 30 | 32 | 35 | 41 | 45 |
|-------|----|----|----|----|----|----|----|----|----|----|----|

**Solution :**

$$D_8 = \frac{N + 1}{10} \times 8$$

$$N = 11$$

$$D_8 = \frac{11+1}{10} \times 8$$

$$D_8 = \frac{12}{10} \times 8 = \frac{96}{10} = 9.6$$

9th item + 0.6 x 10th item - 9th item

35 + 0.6 x 41 - 35

35 + 0.6 x 6

35 + 3.6 = 38.6

$D_8 = 38.6$

## 7.3.2 Discrete Series

$$D = \frac{N+1}{10} \times \text{Required decile}$$

N = Number of Items.

### Illustration 9 :

Calculate 7th decile from the following data.

| Height (cm) | 157 | 168 | 173 | 152 | 162 | 176 |
|---|---|---|---|---|---|---|
| No. of persons | 10 | 13 | 2 | 1 | 25 | 1 |

**Solution :**

Arrange data in order.

| Height (cm) | 152 | 157 | 162 | 168 | 173 | 176 |
|---|---|---|---|---|---|---|
| No. of persons | 1 | 10 | 25 | 13 | 2 | 1 |
| Cumulative Frequency | 1 | 11 | 36 | 49 | 51 | 52 |

$$D_7 = \frac{N+1}{10} \times \text{Required decile}$$

N = 7

$$D_7 = \frac{52+1}{10} \times 7 \qquad = \frac{53}{10} \times 7 = 37.1 \text{ item}$$

$D_7 = 168$

### 7.3.3 Continuous Series

$$d = \frac{N}{10} \times \text{Required Number}$$

$$D = l_1 + \frac{l_2 - l_1}{f_1} \times d - C$$

**Illustration 10 :**

From the following data calculate 6th decile.

| Marks | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 |
|-------|------|-------|-------|-------|-------|-------|-------|-------|
| No. of Students | 5 | 25 | 40 | 70 | 90 | 40 | 20 | 10 |

**Solution**

| Marks | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 |
|-------|------|-------|-------|-------|-------|-------|-------|-------|
| No. of Students | 5 | 25 | 40 | 70 | 90 | 40 | 20 | 10 |
| Cumulative Frequency | 5 | 30 | 70 | 140 | 230 | 270 | 290 | 300 |

$$D = \frac{N}{10} \times 6 \text{ th item, } N = 300$$

$$= \frac{300}{10} \times 6 = 180$$

6th Decile lies in the class 40 - 50

$$D_6 = 40 + \frac{50 - 40}{90} \times 180 - 140$$

$$= 40 + \frac{10}{90} \times 40 = 40 + \frac{400}{90} = 4.44$$

$$D_6 = 44.44$$

### 7.4 PERCENTILES :

Percentiles divide the series into 100 parts. For any series, there are 99 percentiles. Percentiles is denoted by P. It ranges from $P_1$ to $P_{99}$.

### 7.4.1 Individual Series :

$$P = \frac{N + 1}{100} \times \text{Required Percentile}$$

N = No. of Items

**Illustration** 11 :

From the following data calculate 61st percentile.

| Values | 22 | 26 | 14 | 30 | 18 | 11 | 35 | 41 | 12 | 32 |
|--------|----|----|----|----|----|----|----|----|----|----|

**Solution :**

Inorder

| Values | 11 | 12 | 14 | 18 | 22 | 26 | 30 | 32 | 35 | 41 |
|--------|----|----|----|----|----|----|----|----|----|----|

$$P_{61} = \frac{N+1}{100} \times 61$$

N = 10

$$P_{61} = \frac{10+1}{100} \times 61$$

$$P_{61} = \frac{671}{100} = 6.71 \text{ th item}$$

6th item + 0.71 x 7th item - 6th item

26 + 0.71 x (30 - 26)

26 + 0.71 x 4

26 + 2.84 = 28.84

$$P_{61} = 28.84$$

**7.4.2 Discrete Series :**

$$P = \frac{N+1}{100} \times \text{Required Percentile}$$

N = Total of the Frequency

**Illustration 12:**

From the following data calculate 95th Percentile.

| Heights (cms) | 152 | 155 | 157 | 160 | 162 | 164 | 168 | 170 | 171 | 172 | 173 | 176 |
|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| No. of Persons | 1 | 5 | 10 | 12 | 25 | 38 | 13 | 6 | 4 | 3 | 2 | 1 |

**Solution**

| Heights (cms) | 152 | 155 | 157 | 160 | 162 | 164 | 168 | 170 | 171 | 172 | 173 | 176 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| No. of Persons | 1 | 5 | 10 | 12 | 25 | 38 | 13 | 6 | 4 | 3 | 2 | 1 |
| Cumulative Frequency | 1 | 6 | 16 | 28 | 53 | 91 | 104 | 110 | 114 | 117 | 119 | 120 |

$$P_{95} = \frac{N+1}{100} \times 95$$

$$N = 120$$

$$P_{95} = \frac{120+1}{100} \times 95 = 114.95 \text{ th item}$$

$$P_{95} = 172$$

**7.4.3 Continous Series :**

$$p = \frac{N}{100} \times \text{Required Percentile}$$

$$P = l_1 + \frac{l_2 - l_1}{f_1} \times p - C$$

**Illustration 13**

Find out 85th Percentile.

| Marks | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 |
|---|---|---|---|---|---|---|---|---|
| No. of Students | 8 | 12 | 20 | 32 | 30 | 28 | 12 | 4 |

**Solution**

| Marks | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 |
|---|---|---|---|---|---|---|---|---|
| No. of Students | 8 | 12 | 20 | 32 | 30 | 28 | 12 | 4 |
| Cumulative Frequency | 8 | 20 | 40 | 72 | 102 | 130 | 142 | 146 |

$$p = \frac{N}{100} \times 85, \quad N = 146$$

$$p = \frac{146}{100} \times 85 \quad = 124.1 \text{ th item}$$

$$P_{85} = l_1 + \frac{l_2 - l_1}{f_1} \times p - C$$

$$= 50 + \frac{60 - 50}{28} \times 124.1 - 102$$

$$= 50 + \frac{10}{28} \times 22.1 = 50 + 7.89 = 57.89$$

$$P_{85} = 57.89$$

## 7.5  SUMMARY :

Thus, besides Median,there are other positional measures which divide a series into equal parts. Important amongst these are quartiles, deciles and percentiles. In economics and business statistics quartiles are more widely used than deciles and percentiles. The deciles and percentiles are important in psychological and educational statistics concerning grades, ranks and rates etc.

## 7.6  EXERCISE :

1. Explain two quartiles.

2. Explain percentiles.

3. From the following information fingout $Q_1$, $Q_2$, $D_6$, $P_{20}$.

| Marks : | 0 -10 | 10 - 20 | 20 - 30 | 30 - 40 | 40 - 50 | 50 - 60 | 60 - 70 | 70 - 80 |
|---|---|---|---|---|---|---|---|---|
| No. of Students : | 5 | 25 | 40 | 70 | 90 | 40 | 20 | 10 |

(Ans. : $Q_1 = 30.71$, $Q_3 = 49.44$, $D_6 = 44.4$, $P_{20} = 27.5$ )

4. From the following data compute 1st decile, 7th decile, 9th decile, 33rd percentile.

Marks : 35, 76, 63, 24, 12, 95, 47, 55, 85, 93, 3, 18, 29, 59, 69, 30,

29, 51, 68, 71, 80, 99, 8, 13, 41, 89, 73, 20, 9, 5.

(Ans. $D_1 = 8.1$, $D_7 = 70.4$, $D_9 = 92.6$, $P_{33} = 10.23$ )

5. From the following data calculate $Q_1$, $Q_3$, $D_6$, $P_3$.

| Marks | No. of Students |
|---|---|
| Less than 80 | 100 |
| Less than 70 | 90 |
| Less than 60 | 80 |
| Less than 50 | 60 |

| Less than 40 | 32 |
| Less than 30 | 20 |
| Less than 20 | 13 |
| Less than 10 | 5 |

(Ans. : $Q_1 = 34.25$, $D_6 = 50$, $P_3 = 6$)

6. From the following data find out $Q_1$, $Q_3$, $D_2$, $P_{90}$.

| Weight | Below 10 | 10 - 20 | 20 - 40 | 40 - 60 | 60 - 80 | 80 - 100 |
|---|---|---|---|---|---|---|
| No. of Persons | 8 | 10 | 22 | 25 | 10 | 5 |

(Ans. $Q_1 = 21.82$, $Q_3 = 56$, $D_2 = 18$ )

7. From the following data computer $D_7$, $P_{85}$.

| Deposits (Rs.) | 0 - 100 | 100-250 | 250-400 | 400-500 | 500-550 | 550-600 | 600-800 | 800-900 | 900-1000 |
|---|---|---|---|---|---|---|---|---|---|
| No. of Deposits | 25 | 100 | 175 | 74 | 66 | 35 | 5 | 18 | 2 |

(Ans. $D_7 = 467.57$, $P_{85} = 538.64$)

**Dr. K.Kanaka Durga.**

# LESSON - 8
# AVERAGES - IV MODE

**OBJECTIVES :**

After studying this lesson you should be able to understand.

1. What is Mode.

2. How to Calculate Mode

3. What are the merits and limitations

4. Geometric Mean

5. Harmonic Mean

**STRUCTURE OF LESSON :**

8.1 Introduction

8.2 Mode - Definition, Meaning

8.3 Calculation of Mode

    8.3.1   Individual Series

    8.3.2   Discrete Series

    8.3.3   Continuous Series

8.4 Mode with the help of Graph

8.5 Mode - Its Merits and Limitations

8.6 Summary

8.7 Exercise

## 8.1 INTRODUCTION :

Mode like median is also a positional measure. Mode is useful in determining the stock of different goods. Since mode helps us in determining the popularity of a Commodity so it gives opportunity to the business men to stock such items as to get windfall gains.

## 8.2 DEFINITION AND MEANING :

The most frequently occurring item of the series is known as mode. Mode is defined by "Croxton and Cowden" as "The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded as the most typical of a series of values". According to 'Zizek', Mode is "The value occuring most frequently in a series of items and around which the other items are distributed most densely".

The mode is the item which is repeated maximum times in the series will be the mode of the series. Thus in a given series, mode is the most popular and common item. This word is derived from the French word, *Lamode* which means fashion or the most popular phenomenon.

Mode is the most specific or typical value of a series or the value around which maximum concentration of items occur.

For instance. A shirt maker would like to know the size of shirts that has the maximum demand. He will produce shirts of that size, which has the maximum demand.

## 8.3 CALCULATION OF MODE :

Mode can be calculated in Individual series, discrete series and Continuous series.

### 8.3.1 Individual Series :

For degerming mode count the number fo times the various values rpeat themselves and the value which occurs the maximum number of times is the model value.

**Illustration**

Find mode from the following data.

Values : 110    120    130    120    110    140    130    120    140    120

Since the value 120 occurs the maximum numbers of times. i.e., 4. Hence the modal value is 120.

When there are two or more values having the same maximum frequency one cannot say which is the model value and hence mode is said to be defined. Such a series is also known as bimodal or multimodal

**Illustration :**

Find out Mode from the following data.

Income (Rs.) : 610    620    630    620    610    640    630    620    630    640

**Solution :**

| Size of Item | No. of Items it occurs |
|:---:|:---:|
| 610 | 2 |
| 620 | 3 |
| 630 | 3 |
| 640 | 2 |

Here Mode is MUltiple Mode because 620, 630 repeated same number of times.

### 8.3.2 Discrete Series :

In discrete series Mode is located by preparing a 'grouping table' and 'analysis table'.

**a) Grouping Table :** A grouping table has six columns.

1. In column one the maximum frequency is marked.

2. In column two frequencies are grouped in two's.

3. In column three leave the first frequency and then group the remaining in two's.

4. In column four group the frequencies in three's.

5. In column five leave the first frequency and group in three's.

6. In column six leave the first two frequencies and then group the remaining in three's. In each of these take the maximum total and mark it in a circle or by bold type.

**b) Analysis Table** :

After preparing grouping table prepare analysis table, while preparing the grouping table. Put column number on the left hand side and the various probable. Values of mode on the right-hand side. The values against which frequencies are the highest are marked in the grouping table and then entered by means of a bar in the relevant box corresponding to values they represent.

**Illustration :**

From the following data calculate Model wage.

| Daily Wage Rs. : | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|
| No. of Workers : | 8 | 17 | 20 | 22 | 19 | 14 | 10 | 8 | 5 | 3 |

**Solution :**

## Statement of Grouping

| DailyWage X | No.ofWorkers 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 41 | 8 | | | | | |
| 42 | 17 | 25 | | | | |
| 43 | 20 | | 37 | 45 | | |
| 44 | 22 | 42 | | | 59 | 61 |
| 45 | 19 | | 41 | 55 | | |
| 46 | 14 | 33 | | | 43 | 32 |
| 47 | 10 | | 24 | 23 | | |
| 48 | 8 | 18 | | | 16 | |
| 49 | 5 | | 13 | | | |
| 50 | 3 | 8 | | | | |

## Satement of Analysis

| V / F | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | ✓ | | | | | | |
| 2 | | | ✓ | ✓ | ✓ | | | | | |
| 3 | | | | ✓ | ✓ | ✓ | | | | |
| 4 | | | | ✓ | | | | | | |
| 5 | | ✓ | ✓ | ✓ | | | | | | |
| 6 | | | ✓ | ✓ | ✓ | | | | | |
| | | 1 | 3 | 6 | 3 | 1 | | | | |

44 repreated six times, so modal wage : 44

**Illustration :**

Find out Mode from the following data.

| Value | : | 60 | 61 | 62 | 63 | 64 | 65 | 66 |
|---|---|---|---|---|---|---|---|---|
| Frequency | : | 27 | 146 | 435 | 398 | 210 | 128 | 98 |

## Statement of Grouping

| Values | Frequency 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 60 | 27 | | | | | |
| 61 | 146 | 173 | 581 | 608 | 979 | |
| 62 | 435 | 833 | | | | |
| 63 | 398 | | 608 | 736 | | 1043 |
| 64 | 210 | 338 | | | 436 | |
| 65 | 128 | | 226 | | | |
| 66 | 98 | | | | | |

## Satement of Analysis

| V / F | 60 | 61 | 62 | 63 | 64 | 65 | 66 |
|---|---|---|---|---|---|---|---|
| 1 | | | ✓ | | | | |
| 2 | | | ✓ | ✓ | | | |
| 3 | | | | ✓ | ✓ | | |
| 4 | | | | ✓ | ✓ | ✓ | |
| 5 | | ✓ | ✓ | ✓ | | | |
| 6 | | | ✓ | ✓ | ✓ | | |
| | | 1 | 4 | 5 | 3 | 1 | |

## Satement of Analysis

| V / F | 60 | 61 | 62 | 63 | 64 | 65 | 66 |
|---|---|---|---|---|---|---|---|
| 1 | | | ✓ | | | | |
| 2 | | | ✓ | ✓ | | | |
| 3 | | | | ✓ | ✓ | | |
| 4 | | | | ✓ | ✓ | ✓ | |
| 5 | | ✓ | ✓ | ✓ | | | |
| 6 | | | ✓ | ✓ | ✓ | | |
| | | 1 | 4 | 5 | 3 | 1 | |

Hence Mode is 63 because it is repeated 5 times.

**Illustration**

| Weight (Pounds) | : | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 |
|---|---|---|---|---|---|---|---|---|---|---|
| No. of Persons | : | | 4 | 16 | 20 | 18 | 10 | 4 | 25 | 3 | 2 |

**Solution :**

### Statement of Grouping

| Weights | No.ofPersons | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 135 | 4 | | | | | |
| | | 20 | | | | |
| 136 | 16 | | 36 | | | |
| | | | | 40 | | |
| 137 | 20 | | | | 54 | |
| | | 38 | | | | 48 |
| 138 | 18 | | | | | |
| | | | 28 | | | |
| 139 | 10 | | | 32 | | |
| | | 14 | | | 39 | |
| 140 | 4 | | | | | 32 |
| | | | 29 | | | |
| 141 | 25 | | | | | |
| | | 28 | | 30 | | |
| 142 | 3 | | | | | |
| | | | 5 | | | |
| 143 | 2 | | | | | |

**Satement of Analysis**

| F \ V | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | ✓ | | |
| 2 | | | ✓ | ✓ | | | | | |
| 3 | | ✓ | ✓ | | | | | | |
| 4 | ✓ | ✓ | ✓ | | | | | | |
| 5 | ✓ | ✓ | ✓ | | | | | | |
| 6 | | ✓ | ✓ | ✓ | | | | | |
| | 2 | 4 | 5 | 2 | | | 1 | | |

Here Mode is 137 because it repeated 5 times.

### 8.3.3 Continuous Series :

1. By preparing grouping table and analysis table ascetain the model class.

2. Determine the value of mode by applying the following formula.

$$Z = l_1 + \frac{f_1 - f_0}{2 f_1 - f_0 - f_2} \times l_2 - l_1$$

$l_1$ = Model class lower limit

$l_2$ = Model class upper limit

$f_1$ = General frequency of Model class

$f_2$ = General frequency of Succeeding class of Modal class

$f_0$ = General frequency of Preceeeding class of Modal class

**Illustration :**

From the following data calculate Mode.

| Values | 0-5 | 5-10 | 5-15 | 15-20 | 20-25 | 25-30 | 30-35 | 35-40 | 40-45 |
|---|---|---|---|---|---|---|---|---|---|
| Frequency | 20 | 24 | 32 | 28 | 20 | 16 | 37 | 10 | 8 |

## StatementofGrouping

| Values X | Frequence 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 0-5 | 20 | | | | | |
| 5-10 | 24 | 44 | 56 | | | |
| 10-15 | 32 | | | 76 | 84 | |
| 15-20 | 28 | 60 | 48 | | | 80 |
| 20-25 | 20 | | | 64 | | |
| 25-30 | 16 | 36 | 53 | | 73 | |
| 30-35 | 37 | 47 | | 55 | | 63 |
| 35-40 | 10 | | 18 | | | |
| 40-45 | 8 | | | | | |

## SatementofAnalysis

| V / F | 0-5 | 5-10 | 10-15 | 15-20 | 20-25 | 25-30 | 30-35 | 35-40 | 40-45 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | ✓ | | |
| 2 | | | ✓ | ✓ | | | | | |
| 3 | | ✓ | ✓ | | | | | | |
| 4 | ✓ | ✓ | ✓ | | | | | | |
| 5 | | ✓ | ✓ | ✓ | | | | | |
| 6 | | | ✓ | ✓ | ✓ | | | | |
| | 1 | 3 | 5 | 3 | 1 | | 1 | | |

Here Modal class is 10 - 15 because it is repeated 5 times.

Then apply the following principle to find out Mode.

$$Z = l_1 + \frac{f_1 - f_0}{2 f_1 - f_0 - f_2} \times l_2 - l_1$$

$l_1 = 10, l_2 = 15, f_1 = 32, f_2 = 28, f_0 = 24$

$$Z = 10 + \frac{32 - 24}{2 \times 32 - 24 - 28} \times 15\text{-}10$$

$$Z = 10 + \frac{8}{64 - 24 - 28} \times 5$$

$$Z = 10 + \frac{8}{12} \times 5$$

$$Z = 10 + \frac{40}{12} = 10 + 3.33$$

$$Z = 13.33$$

**Inclusive Method**

From the following data calculate Mode.

| Class | 1-5 | 6-10 | 11-15 | 16-20 | 21-25 | 26-30 | 31-35 | 36-40 | 41-45 |
|---|---|---|---|---|---|---|---|---|---|
| Frequency | 7 | 10 | 16 | 32 | 24 | 18 | 10 | 5 | 1 |

**Solution :**

**Statement
of Grouping**

| Values X | Frequency 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 0.5-5.5 | 7 | | | | | |
| 5.5-10.5 | 1026 | | | 1733 | | |
| 10.5-15.5 | 16 | 48 | | | 5872 | |
| 15.5-20.5 | 325674 | | | | | |
| 20.5-25.5 | 2442 | | | | | |
| 25.5-30.5 | 18 | | | | 285233 | |
| 30.5-35.5 | | 106 | 1516 | | | |
| 35.5-40.5 | 5 | | | | | |
| 40.5-45.5 | 1 | | | | | |

**Statement of Analysis**

| F \ V | 0.5-5.5 | 5.5-10.5 | 10.5-15.5 | 15.5-20.5 | 20.5-25.5 | 25.5-30.5 | 30.5-35.5 | 35.5-40.5 | 40.5-45.5 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | | ✓ | | | | | | |
| 2 | | ✓ | ✓ | | | | | | |
| 3 | | | ✓ | ✓ | | | | | |
| 4 | | | ✓ | ✓ | ✓ | | | | |
| 5 | | ✓ | ✓ | | | | | | |
| 6 | ✓ | ✓ | ✓ | ✓ | | | | | |
| | 1 | 3 | 6 | 3 | 1 | | | | |

$$Z = l_1 + \frac{f_1 - f_0}{2 f_1 - f_0 - f_2} \times l_2 - l_1$$

$2f_1 = f_0 = f_2 \; l_1 = 15.5, \; l_2 = 20.5, \; f_1 = 32, \; f_2 = 24, \; f_0 = 16$

$$Z = 15.5 + \frac{32 - 16}{2 \times 32 - 16 - 24} \times 20.5 \square 15.5$$

$$Z = 15.5 + \frac{16}{24} \times 5$$

$$Z = 18.75$$

**Unequal Classes :**

**Illustration**

| Values | 0-2 | 2-4 | 4-8 | 8-10 | 10-15 | 15-20 | 20-25 | 25-30 | 30-35 | 35-40 | 40-50 | 50-60 | 60-70 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| f | 1 | 2 | 2 | 3 | 6 | 8 | 10 | 15 | 18 | 22 | 36 | 10 | 4 |

**Statement of Grouping**

| Values X | Frequency 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 0-10 | 8 | 22 | | 47 | | |
| 10-20 | 14 | | 39 | | | |
| 20-30 | 25 | 65 | | | | |
| 30-40 | 40 | | | | 79 | |
| 40-50 | 36 | 46 | 76 | 86 | | |
| 50-60 | 10 | | | | | |
| 60-70 | 4 | | | | | |
| | | | 14 | | 50 | 104 |

Modal Class = 30-40

**Satement of Analysis**

| F \ V | 10-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|---|---|---|---|---|---|---|---|
| 1 | | | | ✓ | | | |
| 2 | | | ✓ | ✓ | | | |
| 3 | | | | ✓ | ✓ | | |
| 4 | | | | ✓ | ✓ | ✓ | |
| 5 | | | ✓ | ✓ | | | |
| 6 | | ✓ | ✓ | ✓ | ✓ | | |
| | | 1 | 3 | 6 | 3 | 1 | |

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times l_2 - l_1$$

$l_1 = 30,\ l_2 = 40,\ f_1 = 40,\ f_2 = 36,\ f_0 = 25$

$$Z = 30 + \frac{40 - 25}{2 \times 40 - 25 - 36} \times 40 \square 30$$

$$Z = 30 = \frac{15}{19} \times 10$$

$$Z = 30 + \frac{150}{19} = 30 + 7.89$$

Z = 37.89

**Less than - More than**

1. Change the cumulative frequency into general frequency.

2. Construct the class.

**Illustration :**

From the following data calculate Modal mark.

| M arks | N o. of Students |
|---|---|
| Less than 5 | 29 |
| Less than 10 | 224 |
| Less than 15 | 465 |
| Less than 20 | 582 |
| Less than 25 | 634 |
| Less than 30 | 644 |
| Less than 35 | 650 |
| Less than 40 | 653 |
| Less than 45 | 655 |

## Statement of Grouping

| Marks X | No.ofStudents 1 | General Frequence 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 0-5 | 29 | | | | | |
| 5-10 | 195 | 224 | | 465 | | |
| 10-15 | 241 | | 436 | | 553 | |
| 15-20 | 117 | 358 | | | | 410 |
| 20-25 | 52 | | 169 | 179 | | |
| 25-30 | 10 | 62 | | | 68 | |
| 30-35 | 6 | | 16 | | | 19 |
| 35-40 | 3 | 9 | | 11 | | |
| 40-45 | 2 | | 5 | | | |

## Satement of Analysis

| V \\ F | 0-5 | 5-10 | 10-15 | 15-20 | 20-25 | 25-30 | 30-35 | 35-40 | 40-45 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | | ✓ | | | | | | |
| 2 | | | ✓ | ✓ | | | | | |
| 3 | | ✓ | ✓ | | | | | | |
| 4 | ✓ | ✓ | ✓ | | | | | | |
| 5 | | ✓ | ✓ | ✓ | | | | | |
| 6 | | | ✓ | ✓ | ✓ | | | | |
| | 1 | 3 | 6 | 3 | 1 | | | | |

Modal Class - 10 - 15

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times l_2 - l_1$$

$l_1 = 10, l_2 = 15, f_1 = 241, f_2 = 117, f_0 = 195$

$$Z = 10 + \frac{241 - 195}{2 \times 241 - 195 - 117} \times 40 - 30$$

$$Z = 10 + \frac{46}{170} \times 5$$

$Z = 10 + 1.3$

$Z = 11.35$

### Illustration

From the following data calculate Mode.

| Mid Values | Frequency |
|---|---|
| Above 0 | 80 |
| Above 10 | 77 |
| Above 20 | 72 |
| Above 30 | 65 |
| Above 40 | 55 |
| Above 50 | 43 |
| Above 60 | 28 |
| Above 70 | 16 |
| Above 80 | 10 |
| Above 90 | 8 |
| Above 100 | 0 |

**Solution:**

## Statement of Grouping

| Marks X | No.ofStudents 1 | 2 | 3 | 4 | 5 | 6 |
|---------|-----------------|---|---|---|---|---|
| 0-10 | 3 | | | | | |
| 10-20 | 5 | 8 | | 15 | | |
| | | | 12 | | | |
| 20-30 | 7 | | | | 22 | |
| | | 17 | | | | 29 |
| 30-40 | 10 | | | | | |
| | | | 22 | 37 | | |
| 40-50 | 12 | | | | | |
| | | 27 | | | 39 | |
| 50-60 | 15 | | | | | 33 |
| | | | 27 | 20 | | |
| 60-70 | 12 | | | | | |
| | | 18 | | | | |
| 70-80 | 6 | | | | 16 | |
| | | | | | | 10 |
| 80-90 | 2 | | | | | |
| | | 10 | 8 | | | |
| 90-100 | 8 | | | | | |
| 100-110 | 0 | | | | | |

## Satement of Analysis

| V / F | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 | 80-90 | 90-100 |
|-------|------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| 1 | | | | | ✓ | | | | | |
| 2 | | | | ✓ | ✓ | | | | | |
| 3 | | | | | ✓ | ✓ | | | | |
| 4 | | | ✓ | ✓ | ✓ | | | | | |
| 5 | | | | ✓ | ✓ | ✓ | | | | |
| 6 | | | | ✓ | ✓ | ✓ | | | | |
| | | | 1 | 3 | 6 | 3 | 1 | | | |

$$Z = l_1 + \frac{f_1 - f_0}{2 f_1 - f_0 - f_2} \times l_2 - l_1$$

$$Z = 50 + \frac{15 - 12}{2 \times 15 - 12 - 12} \times 60 - 50$$

$$Z = 50 + \dfrac{3}{30 - 12 - 12} \times 10$$

$$Z = 50 + \dfrac{3}{6} \times 10$$

$$Z = 50 + 5$$

$$Z = 55$$

**Illustration**

From the following data calculate Mode.

| Mid Values | 10 | 20 | 30 | 40 | 50 | 60 | 70 |
|---|---|---|---|---|---|---|---|
| Frequency | 7 | 12 | 17 | 29 | 31 | 5 | 3 |

**Solution**

## Statement of Grouping

| Class | Frequency 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 5- | 15 | | | 7 | | |
| | | | | 1936 | | |
| 15-25 | 12 | | 29 | | | |
| 25-35 | 1758 | 46 | | | | |
| 35-45 | 2977 | | | | | |
| | | | | 6065 | | |
| 45-55 | 31 | | | | | |
| | 3639 | | | | | |
| 55-65 | 5 | | 8 | | | |
| 65-75 | 3 | | | | | |

**Satement of Analysis**

| F \ V | 5-15 | 15-25 | 25-35 | 35-45 | 45-55 | 55-65 | 65-75 |
|---|---|---|---|---|---|---|---|
| 1 | | | | ✓ | | | |
| 2 | | ✓ | ✓ | | | | |
| 3 | | | ✓ | ✓ | | | |
| 4 | | | ✓ | ✓ | | ✓ | |
| 5 | ✓ | ✓ | ✓ | | | | |
| 6 | | ✓ | ✓ | ✓ | | | |
| | 1 | 3 | 5 | 4 | | 1 | |

$$Z = 35 + \frac{12}{10} \times 10$$

$$Z = 35 + \frac{12}{10} \times 10$$

$$Z = 35 + 1.2 \times 10 = 35 + 12 = 47$$

$$Z = 47$$

Here Mode should not lies in the Modal class so, Mode can be obtained with the following principle.

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times l_2 - l_1$$

$$Z = 35 + \frac{31}{17 + 31} \times 45 - 35$$

$$Z = 35 + \frac{31}{48} \times 10$$

$$Z = 35 + 6.4$$

$$Z = 41.4$$

**Multiple Mode**

## Statement of Grouping

| Class | Frequency 1 | 2 | 3 | 4 | 5 | 6 |
|-------|-------------|---|---|---|---|---|
| 90-100 | 3 | | | | | |
| | | 5 | | 23 | | |
| 100-110 | 2 | | 20 | | | |
| 110-120 | 18 | | | | 42 | |
| | | 40 | | | | 61 |
| 120-130 | 22 | | 43 | 62 | | |
| 130-140 | 21 | | | | | |
| | | 40 | | | 50 | 32 |
| 140-150 | 19 | | 29 | 15 | | |
| 150-160 | 10 | | | | | |
| | | 13 | | | | |
| 160-170 | 3 | | 5 | | | |
| 170-180 | 2 | | | | | |

## Satement of Analysis

| F \ V | 90-100 | 100-110 | 110-120 | 120-130 | 130-140 | 140-150 | 150-160 | 160-170 | 170-180 |
|-------|--------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1 | | | ✓ | | | | | | |
| 2 | | ✓ | ✓ | ✓ | ✓ | | | | |
| 3 | | | ✓ | ✓ | | | | | |
| 4 | | | ✓ | ✓ | ✓ | | | | |
| 5 | | ✓ | | ✓ | ✓ | ✓ | | | |
| 6 | | | ✓ | ✓ | | | | | |
| | | 2 | 5 | 5 | 3 | 1 | | | |

Here Mode is Multiple. In case of Multiple Mode apply the following principle to locate Mode.

Z = (3 x Median) - (2 x Mean)

**Median :**

| Class | Frequency | cf |
|-------|-----------|-----|
| 90-100 | 3 | 3 |
| 100-110 | 2 | 5 |
| 110-120 | 18 | 23 |
| 120-130 | 22 | 45 |
| 130-140 | 21 | 66 |
| 140-150 | 19 | 85 |
| 150-160 | 10 | 95 |
| 160-170 | 3 | 98 |
| 170-180 | 2 | 100 |

$$\text{Median} = \frac{N}{2} \text{ nd item}$$

$$= \frac{100}{2} \text{ nd item}$$

$$= 130 + \frac{140-130}{21} \text{ x } 50 - 45$$

$$= 130 + \frac{10}{21} \text{ x } 5$$

$$= 130 + \frac{50}{21} \text{ x } 2.38$$

$$\text{Median} = 132.38$$

**Arithmetic Mean :**

| Class | Frequency | Mid Point | dx | fdx |
|-------|-----------|-----------|-----|-----|
| 90-100 | 3 | 95 | -40 | -120 |
| 100-110 | 2 | 105 | -30 | -60 |
| 110-120 | 18 | 115 | -20 | -360 |
| 120-130 | 22 | 125 | -10 | -220 |
| 130-140 | 21 | 135 | 0 | 0 |
| 140-150 | 19 | 145 | +10 | +190 |
| 150-160 | 10 | 155 | +20 | +200 |
| 160-170 | 3 | 165 | +30 | +90 |
| 170-180 | 2 | 175 | +40 | +80 |
|  | 100 |  |  | +560-760 =-200 |

$$= 135 - 2 = 133$$

Z = (3 x Median) - (2 x A.M.)

Median = 132.38

A.M. = 133

Z = (2 x 132.38) - (2 x 133)

Z = 397.14 - 266 = 131.14

Z = 131.14

## 8.4 LOCATING MODE GRAPHICALLY :

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

1. Draw a histogram of the given data.
2. Draw two lines diagonally in the inside of the modal class bar, starting from each upper corner of the bar to the upper corner of the adjacent bar.
3. Draw a perpendicular line from the intersection of the two diagonal lines to the X-axis(horizontal scale) which gives us the modal value.

**Illustration :**

The monthly profits in rupees of 100 shops are distributed as follows.

| Profit per Shop | No. of Shops |
|---|---|
| 0-100 | 12 |
| 100-200 | 18 |
| 200-300 | 27 |
| 300-400 | 20 |
| 400-500 | 17 |
| 500-600 | 06 |

Draw a histogram of the data and find out the Modal value. Check this value by direct calculation.

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times l_2 - l_1$$

Locating ModeGraphically
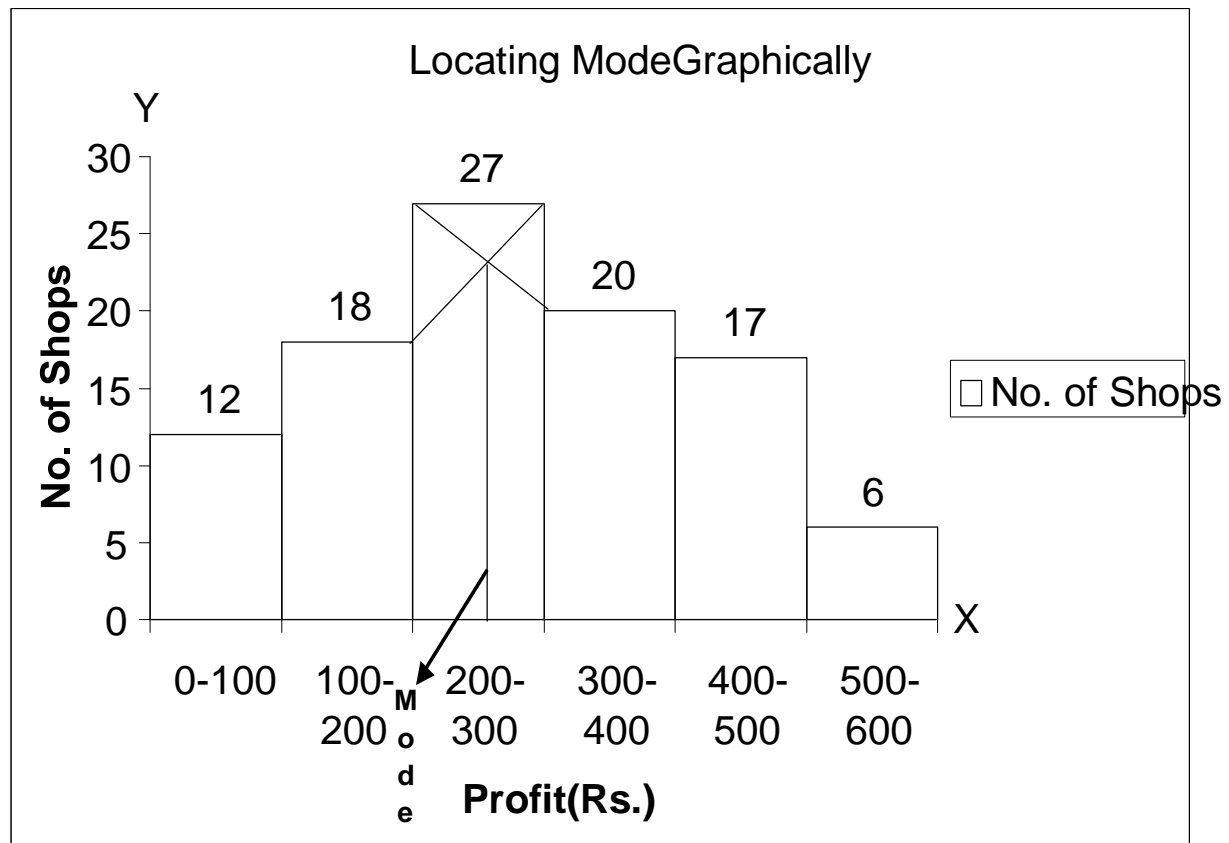
$l_1 = 20, l_2 = 300, f_1 = 27, f_2 = 20, f_0 = 18$

$$Z = 200 + \frac{27 - 18}{2 \times 27 - 18 - 20} \times 300 - 200$$

$$Z = 200 + \frac{9}{54 - 18 - 20} \times 100$$

$$Z = 200 + \frac{900}{16} = 200 + 52.25$$

$$Z = 256.25$$

Mode can also be determined from a frequency polygon in which case a perpendicular is drawn on the base from the apex of the polygon and the point where it intersects the base given the modal value.

Graphic method of determining Mode cannot be determined if two or more classes have the same highest frequency.

## 8.5 MERITS OF MODE :

The main merits of Mode are:

1. The mode is the most frequently occurring value. If the modal wage in a factory is Rs.4100 more than workers receive Rs.4,100 than any other wage. This wage is known as average wage or Modal wage.

2. It is not affected by extremely large or small items.

3. Its value can be determined in open end distributions without ascertaining the class limits.

4. It can be used to describe qualitative phenomenon.

5. The value of Mode can also be determined graphically.

**Limitations :**

The following are the important limitations of Mode.

1. The value of Mode cannot always be determined, because in some cases we may have abimodal series.

2. It is not capable of algebraic manipulations.

3. The value of Mode is not based on each and every item of the series.

4. It is not a rigidly defined measure.

## 8.6 SUMMARY :

Thus the value occurring maximum times is the modal value. This can be known by inspection in Individual Series. In discrete series, mode can be known by inspection. It means to look to that value of the series around which the items are most heavily concentrated. In continuous series after knowing Modal class, grouping and analysis principle is applied to know the Mode.

## 8.7 EXERCISE :

1. Define Mode, How it is useful ?

2. How to locate Mode graphically ?

3. Describe Merits and Limitations of Mode.

4. Calculate Mode.

   49, 35, 21, 46, 57, 67, 57, 13, 99          (Ans. 57)

5. From the following data. Calculate Mode.

   (Ans. 13.5) 6. From the following data calculate.

| Values | : | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|--------|---|---|---|---|---|---|---|---|---|----|----|

| Color Size (in inches) | No. of Persons |
|------------------------|----------------|
| 12.0 | 10 |
| 12.5 | 28 |
| 13.0 | 38 |
| 13.5 | 42 |
| 14.0 | 45 |
| 14.5 | 15 |
| 15.0 | 8 |

15.5                             7

Frequency    :    3    8    10    12    16    14    10    8    17    5

(Ans. : 6)

7. Calculate Mode from the following data.

| Color Size (in inches) | No. of Persons |
|---|---|
| 55 | 8 |
| 65 | 10 |
| 75 | 16 |
| 85 | 14 |
| 95 | 10 |
| 105 | 5 |
| 115 | 2 |

(Ans. 75)

8.  Calculate Mode from the following wages of 50 workers working in a factory.

Daily Wages Rs.:    4    5    7    8    10    11    13    14    16

No. of Workers :    2    3    2    6    10    11    12    3    1

(Ans. : 11)

9.  Find out Mode

Size of Items    : 1   2   3   4   5   6   7   8   9   10   11   12   13   14

Frequency        : 4 10 16 18 24 28 28 30 22 26 18      14   8   6

(Ans. 7)

10.  From the following data findout Modal size.

Values            : 12-13 13-14 14-15 15-16 16-17 17-18 18-19 19-20 20-21 21-22
Frequency    :    5      48      189    303    522    980    981    794    515    474

(Ans. 18.005) 11. Findout Mode.

| Wages | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 | 80-90 |
|---|---|---|---|---|---|---|---|
| No. of Workers | 85 | 120 | 110 | 67 | 49 | 21 | 6 |

(Ans. 37.8)

12. Calculate Mode.

| Class | 20-25 | 25-30 | 30-35 | 35-40 | 40-45 | 45-50 | 50-55 | 55-60 | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|---|
| Frequency | 50 | 70 | 80 | 150 | 180 | 120 | 70 | 50 | (Ans. 42) |

13. Find out Mode.

| Marks | No. of Students |
|-------|-----------------|
| More than 0 | 80 |
| More than 10 | 77 |
| More than 20 | 72 |
| More than 30 | 65 |
| More than 40 | 55 |
| More than 50 | 43 |
| More than 60 | 28 |
| More than 70 | 16 |
| More than 80 | 10 |
| More than 90 | 8 |
| More than 100 | 0 |

**Dr. K. Kanaka Durga**

# LESSON - 9
# AVERAGES : V
## [GEOMETRIC MEAN, HARMONIC MEAN]

**OBJECTIVES :**

After studying this lesson you should be able to understand.

     1. What is Geometric Mean, How to Calculate Geometric Mean

     2. What is Harmonic Mean, How to Calculate Harmonic Mean

**STRUCTURE OF THE LESSON :**

## 9.1 INTRODUCTION :

      There are two means other than Mean, Medin and Mode which are occassionally used in economics and business. These are Geometric Mean and Harmonic Mean. Averages are also called

Ratio - Averages because these are more suitable when the data comprise rates, percentages of ratios insteaded of actual quantities.

## 9.2 GEOMETRIC MEAN (G.M.) - DEFINITON AND MEANING :

Geometric Mean is defined as the nth root of the product of N items or values. If there are two items, we take the square root; if there are three items, the cube root; and so on. Symbolically

Geometric Mean = $^n\sqrt{(X_1)(X_2)......(X_{2n})}$ □ $[(X_1)(X_2)......(X_n)]^{1/n}$

Where $X_1$, $X_2$, $X_n$ refers to the various items of the series.

When the number of items is three or more the task of multiplying the numbers and of exracting the root becomes excessively difficult. To simplify calculations logarithms are used. Geometric Mean is calculated as follows :

Log Geometric Mean = log $X_1$+ log $X_2$+.....log $X_n$

Log Geometric Mean = $\dfrac{\Sigma \log X}{N}$

∴ Geometric Mean = Antilog $\dfrac{\Sigma \log X}{N}$

### 9.2.1 Properties of Geometric Mean :

The following are the important mathematical properties of Geometric Mean.

The product of the value of series will remain unchanged when the value of Geometric Mean is substituted for each individual value.

For Example : The Geometric for series 2, 4, 8 is 4:

Therefore, we have 2 x 4 x 8 = 64 = 4 x 4 x 4

The sum of the deviations of the logarithms of the original observations above or below the logarithms of the geometric mean is equal. This also means that the value of the Geometric mean is such as to balance the ratio deviations of the observations from it.

A note worthy point there is that the Geometric Mean is always lower than arithmetic meanbecuase it gives more weightage to small values. 4. If any item is '0', value of Geometric Mean is also '0'.

### 9.3 CALCULATION OF GEOMETIC MEAN :

Geometric Mena is calculated int he following three Series.

### 9.3.1 Individual Series :

Geometric Mean = Anti log $\dfrac{\Sigma \log X}{N}$

1. Take the logarithms of the variable X and obtain the total $\Sigma \log X$

2. Divide $\Sigma \log X$ by N and take the antilog of the value so obtained.

**Illustration 1:**

From the following data calculate Geometric Mean.

Item : 3, 12, 76, 115, 6, 9, 10, 100, 476, 96

| Sno. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | **10** |
|------|---|---|---|---|---|---|---|---|---|----|--------|
| Item X | 3 | 12 | 76 | 115 | 6 | 9 | 10 | 100 | 476 | 96 | |
| log X | 0.4771 | 1.0792 | 1.8808 | 2.0607 | 0.7782 | 0.9542 | 1.0000 | 2.0000 | 2.6776 | 1.9823 | **14.8909** |

Geometric Mean $= $ Anti log $\dfrac{\Sigma \log X}{N}$

$\Sigma \log X = 14.8909$

$N = 10$

Geometric Mean $=$ Anti log $\dfrac{14.8909}{10}$

Geometric Mean = Anti log of 1.48901

Geometric Mean = 30.83

**Illustration 2 :**

Calculate Geometric Mean.

Values : 85, 70, 15, 75, 500, 8, 45, 250, 40, 36

**Solution**

| Values (X) | 85 | 70 | 15 | 75 | 500 | 8 | 45 | 250 | 40 | 36 | **N = 10** |
|------------|----|----|----|----|-----|---|----|-----|----|----|-----------|
| log X | 1.9294 | 1.8451 | 1.1761 | 1.8751 | 2.6990 | 0.9031 | 1.0532 | 2.3979 | 1.6021 | 1.5563 | **17.6373** |

Geometric Mean $=$ Anti log $\dfrac{\Sigma \log X}{N}$

$\Sigma \log X = 17.6373$  $N = 10$

Geometric Mean $=$ Anti log $\dfrac{17.6373}{10}$

Geometric Mean = Anti log of 1.76373

Geometric Mean = 88.29

**Illustration 3**

  X : 3834, 382, 63, 9, 0.4, 0.03, 0.009, 0.0005

**Solution**

| X | 3834 | 382 | 63 | 9 | 0.4 | 0.03 | 0.009 | 0.0005 | **N = 8** |
|---|------|-----|----|----|-----|------|-------|--------|-----------|
| log X | 3.5837 | 2.5821 | 1.7993 | 0.9031 | 0.3979 | 1.5229 | 2.0458 | 2.3010 | **9.53377** |

$$\text{Geometric Mean} = \text{Anti log} \frac{\Sigma \log X}{N}$$

  Σ log X = 9.53377

   N = 8

$$\text{Geometric Mean} = \text{Anti log} \frac{9.53377}{8}$$

  Geometric Mean = Anti log of 1.9172

  Geometric Mean = 83.60

## 9.3.2 Discrete Series :

$$\text{Geometric Mean} = \text{Anti log} \frac{\Sigma \log xf}{N}$$

1. Find the logarithms of the variable x.

2. Multiplty logarithms with the respective frequencies and obtain the total Σ log x f

3. Divide ☐ log xf by the total frequency and take the anti log of the value so obtained.

**Illustration 4 :**

  From the following wages of 50 workers. Calculate Geometry Mean of Wages.

| Wages | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 |
|-------|----|----|----|----|----|----|----|----|
| No. of Workers | 5 | 7 | 8 | 13 | 9 | 4 | 3 | 1 |

**Solution**

| Wages(x) | No. of Workers | log x | log x f |
|----------|----------------|-------|---------|
| 31 | 5 | 1.4914 | 7.4570 |
| 32 | 7 | 1.5052 | 12.5364 |
| 33 | 8 | 1.5185 | 12.1480 |
| 34 | 13 | 1.5315 | 19.9095 |
| 35 | 9 | 1.5450 | 13.9050 |

| 36 | 4 | 1.5563 | 6.2252 |
|---|---|---|---|
| 37 | 3 | 1.5682 | 4.7046 |
| 38 | 1 | 1.5798 | 1.5798 |
| | **50** | | **76.4655** |

Geometric Mean = Anti log $\dfrac{\Sigma \log xf}{N}$

☐ log xf = 76.4655

N = 50

Geometric Mean = Antilog $\dfrac{76.4655}{50}$

Geometric Mean = Anti log of 1.52931

Geometric Mean = 33.88

**Illustration 5 :**

Calculate Geometric Mean of the following distribution.

| Variable | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|
| Frequency | 11 | 8 | 6 | 9 | 7 | 3 | 1 |

**Solution**

| Variable(x) | Frequency(f) | log x | log x f |
|---|---|---|---|
| 8 | 11 | 0.9031 | 9.9341 |
| 9 | 8 | 0.9542 | 7.6336 |
| 10 | 6 | 1.0000 | 6.0000 |
| 11 | 9 | 1.0414 | 9.3726 |
| 12 | 7 | 1.792 | 1.5544 |
| 13 | 3 | 1.1139 | 3.3417 |
| 14 | 1 | 1.1461 | 1.1461 |
| | **45** | | **44.9825** |

Geometric Mean = Anti log $\dfrac{\Sigma \log xf}{N}$

$\Sigma$ log xf = 44.9825

N = 45

GeometricMean = Anti log $\dfrac{44.9825}{45}$

Geometric Mean = Anti log of 0.9996

Geometric Mean = 9.991

### 9.3.3 Continuous Series :

$$\text{Geometric Mean} = \text{Anti log } \frac{\Sigma \log xf}{N}$$

1.  Find out the Mid point of the classes

2.  Multiply logarithms with the respective frequencies of each class and obtain the total $\Sigma \log xf$ .

3.  Divide the total obtained by the total frequency and take the anti log of the value so obtained.

**Illustration 6 :**

Compute the Geometric Mean from the following data.

| Marks | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 |
|---|---|---|---|---|---|
| No. of Students | 5 | 7 | 15 | 25 | 8 |

Solution

| Marks(x) | No. of Students(f) | Mid Points x | log x | log xf |
|---|---|---|---|---|
| 0-10 | 5 | 5 | 0.6990 | 3.4950 |
| 10-20 | 7 | 15 | 1.1761 | 8.2327 |
| 20-30 | 15 | 25 | 1.3979 | 20.9685 |
| 30-40 | 25 | 35 | 1.5441 | 38.6025 |
| 40-50 | 8 | 45 | 1.6532 | 13.2256 |
|  | **60** |  |  | **84.5243** |

$$\text{Geometric Mean} = \text{Anti log} \frac{\Sigma \log xf}{N}$$

$\Sigma \log xf = 84.5243N$

$N = 60$

$$\text{GeometricMean} = \text{Antilog} \frac{84.5243}{60}$$

Geometric Mean = Anti log of 1.4087

Geometric Mean = 25.63

**Illustration 7**

From the following data. Calculate Geometric Mean.

| Class | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|---|---|---|---|---|---|---|---|
| Frequency | 4 | 8 | 9 | 13 | 7 | 6 | 3 |

**Solution :**

| Class (x) | No. of Students(f) | Mid Points x | log x | log xf |
|-----------|-------------------|--------------|-------|--------|
| 0-10 | 4 | 5 | 0.6990 | 2.7960 |
| 10-20 | 8 | 15 | 1.1761 | 9.4088 |
| 20-30 | 9 | 25 | 1.3976 | 12.5811 |
| 30-40 | 13 | 35 | 1.5450 | 20.0850 |
| 40-50 | 7 | 45 | 1.6532 | 11.5724 |
| 50-60 | 6 | 55 | 1.7404 | 10.4424 |
| 60-70 | 3 | 65 | 1.8129 | 5.4387 |
| | **50** | | | **72.3244** |

$$\text{Geometric Mean} = \text{Anti log} \frac{\Sigma \log xf}{N}$$

$\Sigma \log xf = 72.3244N$

$N = 50$

$$\text{GeometricMean} \square \text{Antilog} \frac{72.3244}{50}$$

Geometric Mean = Anti log of 1.446488

Geometric Mean = 27.957

## 9.4  MERITS AND LIMITATIONS OF GEOMETRIC MEAN :

**Merits or Advantages :**

1.  Geometric Mean is rigidly defined.

2.  It is based on all observations.

3.  It is suitable for further mathematical treatment.

4.  Geometric is useful in fixation of prices etc.

5.  it is not affected much by fluctuations of sampling.

6.  Geometric mean is useful when data are in rates, ratios, percentages, etc.

7.  It is useful for finding the compound rates of change.

8.  It is used in the construction of Index numbers.

**Demerits or Limitations :**

1.  Geometric is not easy to understand and calculate.

2.  If any item is zero, Geometric Mean becomes zero.

## 9.5 HARMONIC MEAN : DEFINITION AND MEANING :

Harmonic Mean is used in Special types of problems. it is based on arithmetic mean or reciprocals of the values of the variable.

Harmonic Mean is defined as the " reciprocals of the values of the variables".

SymbolicallyHarmonic Mean $n / [(1/x_1)+(1/x_2)+(1/x_3)+\ldots+(1/x_n)]$

X = Variable

Harmonic Mean is always less than not only the arithmetic mean but geometric mean as well. It is becuase that the average gives weightage to smaller items i.e. reciprocal of 3 is $\frac{1}{3}$ and that of 4 is $\frac{1}{4}$ . Also number of value of the variable be zero.

## 9.6 CALCULATION OF HARMONIC MEAN :

When the number of items is large the computation of Harmonic Mean is tedious. To simplify calculations we obtain reciprocal of the various items from the table and apply the Principle.

### 9.6.1 Harmonic Mean - Individual Series :

Harmonic Mean $= \dfrac{N}{\sum(1/X)}$

1. Obtain reciprocals of given number.

2. Obtain arithmetic mean of the reciprocals.

3. Find the reciprocal of the arithmetic mean.

**Illustration 8**

Calculate Harmonic Mean

X   1238   178.7   89.9   78.4   9.7   0.989   0.874   0.012   0.008   0.0009

**Solution**

| X | 1 / X |
|---|---|
| 1238.0 | 0.0008 |
| 178.7 | 0.0056 |
| 89.9 | 0.0111 |
| 78.4 | 0.0128 |
| 9.7 | 0.1031 |
| 0.989 | 1.0111 |
| 0.874 | 1.1442 |
| 0.12 | 83.3333 |
| 0.008 | 125.0000 |
| 0.0009 | 1111.1111 |
| **N = 10** | **1321.7331** |

Harmonic Mean $= \dfrac{N}{\sum(1/X)}$

N = 10

$\sum(1/X) + = 1321.7331$

Harmonic Mean $= \dfrac{10}{1321.7331}$

Harmonic Mean = 0.0076

**Illustration 9 :**

Calculate Harmonic Mean.

X 3834 382 63 8 0.4 0.03 0.009 0.0005

**Solution**

| X | 1 / X |
|---|---|
| 3874 | 0.0003 |
| 382 | 0.0027 |
| 63 | 0.0159 |
| 8 | 0.1250 |
| 0.4 | 2.5000 |
| 0.03 | 33.3333 |
| 0.009 | 11.1111 |
| 0.0005 | 2000.0000 |
| **N = 8** | **2147.0883** |

Harmonic Mean $= \dfrac{N}{\sum(1/X)}$

N = 8

$\sum(1/X) = 2147.0883$

Harmonic Mean $= \dfrac{8}{2147.0883}$

Harmonic Mean = 0.003726

### 9.6.2 Harmonic Mean - Discrete Series :

Harmonic Mean $= \dfrac{N}{\sum(1/Xf)}$

1. Take the reciprocal of the various items.

2. Multiply the reciprocals by respective frequencies.

3. Substitute the vlues of N and $\sum(1/Xf)$

Instead of finding out the reciprocals first and then multiplying them by frequencies it will be far more easier to divide each frequency by the respective value of the variable.

Harmonic Mean $= \dfrac{N}{\sum(f/x)}$

**Illustration**

From the following data compute the value of Harmonic mean.

| Marks | 10 | 20 | 25 | 40 | 50 |
|---|---|---|---|---|---|
| No. of Students | 20 | 30 | 50 | 15 | 5 |

**Solution**

| Marks (X) | No. of Students (f) | f / X |
|---|---|---|
| 10 | 20 | 2.000 |
| 20 | 30 | 1.500 |
| 25 | 50 | 2.000 |
| 40 | 15 | 0.375 |
| 50 | 5 | 0.100 |
| | **120** | **5.975** |

Harmonic Mean $= \dfrac{N}{\sum(f/x)}$

$N = 120$

$\sum(f/x) = 5.975$

Harmonic Mean $= \dfrac{120}{5.975}$

Harmonic Mean $= 20.08$

### 9.6.3 Harmonic Mean - Continous Series :

Take the mid points of class.

1. Take the reciprocals of the mid points.

2. Multiply the reciprocals by respective frequencies.

3. Substitute the values in Principle.

**Illustration:**

Calculate Harmonic Mean from the following data.

| Class | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 |
|-------|------|-------|-------|-------|-------|-------|-------|-------|
| Frequency | 8 | 12 | 13 | 15 | 17 | 16 | 14 | 5 |

**Solution :**

| Class (X) | Frequency (f) | Mid Points | (1 / X) | 1 x f X |
|-----------|---------------|------------|---------|---------|
| 0-10 | 8 | 5 | 0.2000 | 1.0000 |
| 10-20 | 12 | 15 | 0.0667 | 1.0005 |
| 20-30 | 13 | 25 | 0.0400 | 1.0000 |
| 30-40 | 15 | 35 | 0.0286 | 1.0010 |
| 40-50 | 17 | 45 | 0.0222 | 0.9990 |
| 50-60 | 16 | 55 | 0.0182 | 1.0010 |
| 60-70 | 14 | 65 | 0.0154 | 0.8645 |
| 70-80 | 5 | 75 | 0.0133 | 0.9975 |
|  | **N = 100** |  |  | **7.8635** |

$$\text{HarmonicMean} \quad \frac{100}{8635} \quad 7.$$

Harmonic Mean = 12.72

## 9.7  MERITS AND LIMITATIONS OF HARMONIC MEAN :

**Merits :**

1. Harmonic Mean is rigidly defined.

2. It is based on all the items.

3. It is suitable for further algebraic treatment.

4. It gives greater weightage to smaller values. (because of reciprocal usage)

5. It is not affected by fluctuations of sampling.

6. It is useful in averaging special types of rates and ratios.

**Demerits :**

1. It is not easy to calculate and understand.

2. If one of the items is zero Harmonic Mean can not be calculated.

3. It is hardly used in business problems. Because it is not a representative figure of the distribution unless the phenomenon needs greater importance to be given to smaller items.

## 9.8 SUMMARY :

Thus the Geometric Mean and Harmonic Mean are two means which are occassionally used in economics and business. These are more suitable when the data comprises rates, percentages of ratios instead of actual quantities. Geometric Mean is also useful for finding the compound rates of chzanges like the rates of growth of population in a country over a period of time or the average rate of increase or decrease in the turnover of a business. Harmonic Mean would be representative when different rates of speed, for equal distances have to be averaged.

## 9.9 EXERCISE :

1. What is Geometric Mean, what are its Merits and limitations.

2. What is Harmonic Mean.

3. When do we use Harmonic Mean.

4. What are the Merits and Demerits of Harmonic Mean.

5. Find the Harmonic Mean 2574, 46575.5, 0.8, 0.08, 0.005, 0.0009 (Ans. : 0.00604)zz
       X 85, 70, 15, 75, 500, 8, 45, 250, 40, 36 7. Calculate        (Ans. : 58.03)
Geometric Mean of the following data.

       X        0.009, 0.005, 0.08, 0.8, 5, 75, 475, 2574        (Ans. : 1.841)

8. Find out Geometric Mean

       X        10, 110, 135, 120, 50, 59, 60, 7                (Ans. : 46.56)

9.   Find the Geometric mean from the following data.

| X | 2 | 3 | 5 | 6 | 4 | |
|---|---|---|---|---|---|---|
| f | 10 | 15 | 18 | 12 | 7 | (Ans.: 3.850) |

10. Compute Geometric Mean

       X        10 15 18 20 25 f 2 3 5 6 4 (Ans.: 18.2)

11. Calculate Geometric Mean

| X | 5 | 15 | 25 | 35 | 45 | |
|---|---|---|---|---|---|---|
| f | 5 | 7 | 15 | 25 | 8 | (Ans. : 25.63) |

12. Find out Geometric Mean

| X | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | |
|---|---|---|---|---|---|---|
| f | 5 | 10 | 15 | 7 | 4 | (Ans. : 31.72) |

13. Compute the Geometric Mean

| X | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 |
|---|---|---|---|---|---|
| f | 10 | 5 | 8 | 7 | 20 |

14. The following marks are related to 60 students in Economics, Compute Geometric Mean.

| Marks | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | |
|---|---|---|---|---|---|---|---|
| No. of Students | 3 | 8 | 15 | 20 | 10 | 4 | (Ans. : 28.02) |

15. Calculate Harmonic Mean.

| X | 10 | 20 | 40 | 60 | 120 | (Ans. : 25) |
|---|---|---|---|---|---|---|

16. Find out Harmonic Mean.

| X | 3834 | 382 | 63 | 0.8 | 0.4 | 0.03 | 0.009 | 0.0005 |
|---|---|---|---|---|---|---|---|---|

(Ans. : 0.00373)

17. Find out Geometric Mean

| X | 10 | 20 | 25 | 40 | 50 | |
|---|---|---|---|---|---|---|
| f | 20 | 30 | 50 | 15 | 5 | (Ans. : 20.08) |

18. Find out Harmonic Mean

| Marks | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | |
|---|---|---|---|---|---|---|
| No. of Sudents | 4 | 6 | 10 | 7 | 3 | (Ans. : 16.03) |

19. Find out Geometric Mean

| Class Interval | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 |
|---|---|---|---|---|---|
| Frequency | 4 | 6 | 10 | 7 | 3 |

**- Dr.K. Kanaka Durga**

# LESSON - 10
# MEASURES OF DISPERSION - I
## (RANGE, QUARTILE DEVIATION &MEAN DEVIATION)

**OBJECTIVES :**

     By the study of this chapter you will be able to understand the meaning of dispersion and three measures of dispersion (Range , Quartile Deviation & Mean Deviation ). You will also be thorough with merits, demerits and method of computing all these three measures of dispersion.

**STRUCTURE OF THE LESSON :**

## 10.1  INTRODUCTION :

     " Measures of Dispersion " or " Measures of Variation " are the "Average of second order" They are based on the average of deviations of the values obtained from the central tendencies i.e. Arithmetic Mean (a), Median (M), or Mode (z). The variability is the basic feature of the values of variables. Such type of variation or dispersion refers to the "lack of uniformity ".

## 10.2 DIFFERENCES BETWEENCENTRAL TENDENCIES AND DISPERSION :

  Following are the distinctions between the central tendencies and Dispersions –

| **Central Tendency** | **Dispersions** |
|---|---|
| 1. Average of the first order | 1. Average of the second order |
| 2. Do not throw light on the formation of series | 2. Throw light on the fromation of series or distribution |
| 3. Do not give detailed features of obseravations | 3. Give detailed characteristics of obseravations |
| 4. Do not establish relationship with the items | 4. Establish relationship with the individual items |
| 5. Do not reveal entire picture of distribution | 5. Reveal the entire picture of the distribution |
| 6. Give only the idea of concentration of item | 6. Give the idea of deviation from central tendencies. |

An average of second order is an average of the difference of all the items of the series from an average of those items. In averaging these differences or deviations, their irregularities are brushed off and a representative of dispersion results in.

All the distributions are not similar. They differ in numerical size of their average and in their respective formations. Let us observe the following series carefully

Series 1 : 30     30     30     30     30     30     30     30     30

Series 2 : 22     24     26     28     30     32     34     36     38

Series 3 :   6     12     18     24     30     36     42     48     54

Arithmetic mean and median in all the series are same i.e. : 30 but items in series differ widely. So the central tendencies fail to describe the scatterdness of the values. For measuring the nature of formation we require the average of second order in support of the first order.

## 10.3 OBJECTIVES :

The objectives of computing the second order averages are given below.

a) To ascertain the suitability of the first order averages

b) To decide the consistency of performance and

c) To reveal the degree of uniformity in the series

In the three series as given above, constituted differently though their mean and median are the same. The first series is uniformly distributed and there is no dispersion at all. The second series is having same sort of dispersion from the central tendency and the uniformity is disturbed.

The third series shows a high degree of dispersion and there is no uniformity among the items. Thus it can be concluded that the larger the dispersion is, the lower will be the uniformity in the distribution.

## 10.4 TYPES OF MEASURES OF DISPERSION :

Measures of dispersion are mainly 4 types -

1.  Range

2.  Quartile Deviation or semi- inter quartile range

3.  Mean deviation or Average deviation

4.  Standard deviation

**1. Range :**

The difference between line Highest value (H) and least value (L) of a series is called the 'Range'. ' Range ' represents the difference between the extreme values. The values in between the two extremes are not at all taken into consideration.

Range (R) = H – L

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

H = Highest value   L = Least value.

**Individual series - Range :** Range (R) = H – L

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

**Example 1 :**

Compute the range and the coefficient of range of the series and state which one is more dispersed and which one is more uniform.

Values of Variables :

|  |  |  |  |  |  |
|---|---|---|---|---|---|
| Series 1 : 13 | 14 | 15 | 16 | 17 | (a =15) |
| Series 2 :  9 | 12 | 15 | 18 | 21 | (a=15) |
| Series 3 :  1 | 8 | 15 | 22 | 29 | (a=15) |

" Central tendency is same but formation differ "

**Solution :**

|  | I | II | III |
|---|---|---|---|
| Range : | R = H – L | R = H – L | R = H – L |
|  | = 17 - 13 | = 21-9 | = 29-1 |
|  | = 4 | = 12 | = 28 |

Coefficient of Range:

$$\frac{H-L}{H+L} \qquad \frac{H-L}{H+L} \qquad \frac{H-L}{H+L}$$

$$\frac{17-13}{17+13} \qquad \frac{21-9}{21+9} \qquad \frac{29-1}{29+1}$$

$$= 1.33 \qquad\qquad = 0.4 \qquad = 0.93$$

Series I is ' Less ' dispersed and more uniform. Series II is ' Less ' Uniform and more dispersed

**Discrete Series :**

Range (R) = H – L

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

        **Note :** The frequencies are not to be taken into consideration in the computation of Range.

**Example 2 :**

From the following distribution find out the Range and its coefficient Values of Variables :

| Marks (x) : | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| No.of Students (f) : | 4 | 7 | 12 | 13 | 18 | 16 | 14 | 9 | 5 | 2 |

**Solution :**

R = H – L

= 10 - 1

= 9

Coefficient of Range : $\dfrac{H - L}{H + L} = \dfrac{9}{11} = 0.81$

**Continuous Series :**

Range (R) = H – L

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

     In finding out the Range in continuous series the frequencies are never taken into account. The upper limit of the Highest class (H) and lower limit of the least class (L) are only taken into account.

**Example 3**

Compute the Range and Coeffient from the following data

      (x) :    10-12    12-14  14-16  16-18    18-20    20-22    22-24  24-26  26-2828-30

| (f) : | 12 | 13 | 18 | 21 | 23 | 27 | 19 | 14 | 11 | 9 |

**Solution :**

$R = H - L$

$= 30\text{-}10$

$= 20$

Coefficient of Range : $\dfrac{H - L}{H + L}$

$= \dfrac{30 - 10}{30 + 10}$

$= \dfrac{20}{40} = 0.5$

## 10.6 MERITS OF RANGE :

Following are the merits of Range.

a) It is the simplest measure of dispersion

b) It is regidly defined and easiest measure of dispersion to compute

c) It is readily comprehensible and it requires very little calculations.

d) It is useful in statictical methods of quality control techniques

e) It is useful in studying the variations in the prices of share and stocks.

f) It is useful in studying weather conditions ( weatheriology or meterology ) where minimum and maximum temperature is identified

## 10.7 DEMERITS OF RANGE :

a) Unfortunately it is not a stable measure of dispersion, because it is affected by the extreme values only.

b) It is not suitable where the class intervals are open in the distribution.

c) It is completely depending upon the two extreme values but not on the other values.

d) It is not suitable for mathematical treatment

e) It is very sensitive to the fluctuations in the sampling size as the size of sample increase it tends to increase not in proportion.

## 10.8 QUARTILE DEVIATION OR SEMI - INTER QUARTILE RANGE :

**Introduction :** One of the demerits of the Range is that it is only affected by the extreme values. To over come this defect, the Quartile deviation is formulated with some modifications. It is similar to Range. For the computation of Quartile Deviation $Q_3$ and $Q_1$ will be taken as Highest and

Least values. It means, the items below the Lower Quartile ($Q_1$) and the items above the Upper Quartile ($Q_3$) are not considered. It means only the middle may portion of the series will be considered. The range so obtained is divided by two to get the Quartile Deviation. Thus the Quartile Deviation measures the difference between the values of $Q_1$ and $Q_3$.

$$\text{Quartile Deviation} = Q.D = \frac{Q_3 - Q_1}{2} \qquad \text{Absolute Measure 2}$$

$$\text{Coefficient of Range} = \frac{Q_3 - Q_1}{Q_3 + Q_1} \text{--------- Relative measure.}$$

Where $Q_1$ means first Quartile or Lower Quartile

   $Q_3$ means Third Quartile or Upper Quartile

**Individual series:**

$$Q.D = \frac{Q_3 - Q_1}{2}$$

$$\frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$Q_3 = \frac{N+1}{4} \text{ x 3rd item.}$$

**Note :** The series must be arranged in an ascending order.

**Example 4**

From the following data, compute quartile Deviation and Co-efficient of Quartile Deviation.
Variables : 24,7,11,9,17,3,20,14,4,22,27

**Solution :**

Arranging the series in Ascending order :

| S.No. | X |
|---|---|
| 1 | 3 |
| 2 | 4 |
| 3 | 7 |
| 4 | 9 |
| 5 | 11 |
| 6 | 14 |
| 7 | 17 |
| 8 | 20 |
| 9 | 22 |

10      24

11      <u>27</u>

<u>          N:11</u>

$$Q_1 = \frac{N+1}{4} \text{ th item}$$

$$= \frac{11+1}{4} \text{ th item} = \frac{12}{4} \text{ th item} = \text{3rd item}$$

3rd item = 7

$$Q_3 = \frac{N+1}{4} \text{ x 3rd item}$$

$$= \text{x 3rd item } \frac{11+1}{4} = \text{x3rd item } \frac{12}{4}$$

9th item = 22

$$Q.D = \frac{Q_3 - Q_1}{2}$$

$$= \frac{22 - 7}{2}$$

$$\text{Coefficient of Range} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$\frac{22 - 7}{22 + 7}$$

$$\frac{15}{29} = 0.52$$

**Example 5 :**

From the following Marks of 12 students compute the Quartile Deviation and its coefficient.
Marks : 43, 54, 67,80,89,84,72,61,48,30,25,37

**Solution :**

Ascending order -

| S.No. | X |
|---|---|
| 1 | 25 |
| 2 | 30 |
| 3 | 37 |
| 4 | 43 |
| 5 | 48 |
| 6 | 54 |
| 7 | 61 |
| 8 | 67 |
| 9 | 72 |
| 10 | 80 |
| 11 | 89 |

N:12

$$Q_1 = \frac{N+1}{4} \text{th item}$$

$= \frac{12+1}{4}$ th item = 3.25 th item = 3rd item + 25% of 6 ( 43-37)

37+1.5 = 38.5

$$Q_3 = \frac{N+1}{4} \times 3\text{rd item}$$

$= \frac{12+1}{4} \times 3\text{rd item}$

9.75th item = 9th item + 75% of 8 ( 80-72)

$= 72 + 6 = 78$

$$Q.D = \frac{Q_3 - Q_1}{2}$$

$= \frac{78 - 38.5}{2} = 19.75$

Coefficient of Range $= \dfrac{Q_3 - Q_1}{Q_3 - Q_1}$

$$= \frac{\overline{78 - 38.5}}{78 + 38.5}$$

$$= \frac{\overline{39.5}}{116.5} = 0.339$$

**Discrete Series - Quartile Deviation :**

$$Q.D = \frac{Q_3 - Q_1}{2}$$

$$\text{Coefficient of Range} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$Q_1 = \frac{N + 1}{4} \text{ th item}$$

$$Q_3 = \frac{N + 1}{4} \text{ x 3rd item.}$$

These two items must be identified in the ' Cf ' and the corresponding variables shall be taken as $Q_1$ and $Q_3$

**Example 6 :**

compute the Quartile Deviation and its coefficient from the following data.

| (x) : | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|-------|----|----|----|----|----|----|----|----|----|----|
| (f) : | 4  | 8  | 13 | 16 | 18 | 14 | 11 | 9  | 5  | 2  |

**Solution :**

| X | f | cf |
|----|----|-----|
| 21 | 4 | 4 |
| 22 | 8 | 12 |
| 23 | 13 | 25 |
| 24 | 16 | 41 |
| 25 | 18 | 59 |
| 26 | 14 | 73 |
| 27 | 11 | 84 |
| 28 | 9 | 93 |
| 29 | 5 | 98 |
| 30 | 2 | 100 |

N:100

$$Q_1 = \frac{N+1}{4}\text{th item}$$

$$= \frac{100+1}{4}\text{ th item} = 25.25\text{ th item}$$

It lies in the cf 41 and the corresponding variable is 24

$$Q_1 = 24$$

$$Q_3 = \frac{N+1}{4}\text{x 3rd item}$$

$$= \frac{100+1}{4}\text{ x 3rd item} = 75.75\text{ th item}$$

It lies in the cf 84 and the corresponding variable is 27

$$Q_3 = 27$$

$$Q.D = \frac{Q3- Q1}{2}$$

$$= \frac{24-27}{2} = \frac{3}{2} = 1.5$$

$$= \frac{24-27}{27+24}$$

$$= \frac{3}{51} = 0.0588$$

**Continuous Series - Quartile Deviation :**

$$Q.D = \frac{Q3-Q1}{2}$$

These items must be identified in the ' cf ' and the corresponding classes shall the taken as $Q_1$ class $Q_3$ class. Then the following formula shall be applied to find the $Q_1$ and $Q_3$

$$l + \frac{c \text{ x } i}{f}$$

Where    l = Lower limit of the class

c = difference between $\dfrac{N}{4}$ th item and the preceding cf

i = interval of the Quartile class f = frequency of the quartile class

**Note :** $Q_1$ and $Q_3$ should not be calculated from the inclusive classes. They must be converted into Exclusive classes.

## Example 7

compute the Quartile Deviation and its coefficient from the following data.

| (x) : | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 | 80-90 | 90-100 |
|-------|------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| (f) : | 4 | 9 | 13 | 15 | 14 | 16 | 12 | 8 | 6 | 3 |

**Solution :**

| X | f | cf |
|------|------|------|
| 0-10 | 4 | 4 |
| 10-20 | 9 | 13 |
| 20-30 | 13 | 26 |
| 30-40 | 15 | 41 |
| 40-50 | 14 | 55 |
| 50-60 | 16 | 71 |
| 60-70 | 12 | 83 |
| 70-80 | 8 | 91 |
| 80-90 | 6 | 97 |
| 90-100 | 3 | 100 |

N= 100

$Q_1$ **Position**    =    $\dfrac{N}{4}$ th item 4

$= \dfrac{100}{4}$ th item = 25 th item

It lies in the cf 26 and corresponding class is 20-30

Thus $Q_1$ class = 20 - 30

$= 20 + \dfrac{(25-13)\text{x}10}{13}$

$$= 20 + 9.232 \quad = 29.232$$

$$\mathbf{Q_3\,Position} \quad = \frac{N}{4} \times 3\text{rd item}$$

$$= \frac{100}{4} \times 3\text{rd item} = 75\text{ th item}$$

H lies in the cf 83 and corresponding class is 60-70

Q$_3$ class = 60-70

$$Q_3 = 1 + \frac{c \times i}{f}$$

$$= 60 + \frac{(75 - 71) \times 10}{12}$$

$$= 60 + 3.33 = 63.33$$

$$\text{Q.D} = \frac{Q3 - Q1}{2}$$

$$= \frac{63.33 - 29.33}{2} = 17.05$$

$$\text{Coefficient of Range} = \frac{Q3 - Q1}{Q3 + Q1}$$

$$= \frac{63.33 - 29.33}{63.33 + 29.33}$$

$$= \frac{34.10}{92.56} = 0.3684$$

**Example 8**

compute the Quartile Deviation and its coefficient from the following data.

| (x) : | 0-9 | 10-19 | 20-29 | 30-39 | 40-49 | 50-59 | 60-69 | 70-79 | 80-89 | 90-99 |
|-------|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| (f) : | 3 | 8 | 12 | 13 | 19 | 18 | 17 | 14 | 9 | 5 |

**Solution :**

| X | f | cf |
|------|----|----|
| 0 - 9 | 3 | 3 |
| 10-19 | 8 | 11 |
| 20-29 | 12 | 23 |

|       |    |     |
|-------|----|-----|
| 30-39 | 13 | 36  |
| 40-49 | 19 | 55  |
| 50-59 | 18 | 73  |
| 60-69 | 17 | 90  |
| 70-79 | 14 | 104 |
| 80-89 | 9  | 113 |
| 90-99 | 5  | 118 |

_____ N= 118

**$Q_1$ Position** = N/4 th item

$$= \frac{118}{4} \text{ th item} = 29.5 \text{ th item}$$

It lies in the cf 36 and corresponding class is 30-39

Thus $Q_1$ class = 30-39 But it is an inclusive class. It must be converted into an exclusive class

Exclusive class = 29.5 - 39.5

$$Q_1 = 1 + \frac{c \times i}{f}$$

$$= 29.5 + \frac{(29.5 - 23) \times 10}{13}$$

$$= 29.5 + 5$$

$$= 34.5$$

**$Q_3$ Position** = N/4 x 3rd item

$$= \frac{118}{4} \text{ x 3rd item} = 88.5 \text{ th item}$$

It lies in the cf 90. The corresponding class is 60-69

$Q_3$ class ( Exclusive form )   = 59.5 - 69.5

$$Q_3 = 1 + \frac{c \times i}{f}$$

$$= 59.5 + \frac{(88.5 - 73) \times 10}{17}$$

$$= 59.5 + 9.12$$

= 68.62

$$Q.D = \frac{Q3-Q1}{2}$$

$$= \frac{68.62 - 34.5}{2} = \frac{34.12}{2}$$

= 17.06

Coefficient of Range = $\frac{Q3-Q1}{Q3+Q1}$

$$= \frac{68.62 - 34.5}{68.62 + 34.5}$$

$$= \frac{34.12}{103.12} = 0.3311$$

## 10. 9 MERITS OF QUARTILE DEVIATION :

1. It is very easy to calculate and simple to understand.

2. It is not affected by extreme values of variable as it is concerned with the central half portion of distribution

3. It is not at all affected by open end class intervals.

## 10.10  DEMERITS OF QUARTILE DEVIATION :

1. It ignores completely the portion below the lower quartile and above the upper quartile

2. It is not capable of further mathematical treatment

3. It is greatly affected by the fluctuation in the sampling

4. It is only a positional average but not mathematical average.

## 10.11  MEAN DEVIATION :

Introduction : The average of deviations taken from an average is called Mean Deviation(M.D) or Average Deviation. The base average may be either Mean or Median or Mode. But theoretically, the deviations of items are taken preferably from median instead that than form the Mean or the Mode. Mediam is supposed to be the suitable central tendency for calculating deviations because the sum of the deviations from the Median is less than the sum of deviations from the Mean. It is not a common pracitce to calculate the deviation from the mode as its value is sometimes not clearly defined.

In aggregating the deviations the algebric negative signs are not taken into account. It means all the deviations are treated as Positive ignoring the negative signs.

**Individual series - Mean Deviation :**

$$M.D = \frac{\in|dx|}{N}$$

Where MD = Mean Deviation

∈|dx| = Total of the deviation taken from the average by ignoring the signs (+or −) N= Number of variables.

first of all an average shall be calculated. It may be either mean or Median.

$$\text{Arithmetic Mean (a)} = \frac{\in x}{N}$$

$$\text{Median} = \frac{N+1}{} \text{ nd item ( after arranging the series in an ascending order )}$$

Deviations must be taken from the average to the other variables in the series by ignoring

plus and minus.The total of these deviations must be devided with the number of deviations $\frac{\in|dx|}{N}$

$$\text{Coefficient of Mean Deviation} = \frac{M.D}{Average}$$

**Example 9**

Find out the Mean deviation from Mean and Median and also find out the coefficient.

(x) :  21  34  27  35  30  24  29  22  33  25

**Solution :** Computation of Mean Deviation from Mean

| S.No. | X | \|dx\| |
|---|---|---|
| 1 | 21 | 7 |
| 2 | 34 | 6 |
| 3 | 27 | 1 |
| 4 | 35 | 7 |
| 5 | 30 | 2 |
| 6 | 24 | 4 |
| 7 | 29 | 1 |
| 8 | 22 | 6 |
| 9 | 33 | 5 |

|  |  |  |
|---|---|---|
| 10 | 25 | 3 |

$$N=10 \qquad N= 280 \qquad 42$$

No. = 10

$\in x = 280$

$$a = \frac{x}{N} = \frac{280}{10} = 28$$

$\in dx = 42$

$$M.D = \frac{\in | dx |}{N} = \frac{42}{10} = 4.2$$

$$\text{Coefficient} = \frac{M.D}{Average} = \frac{4.2}{28} = 0.15$$

Computation of M.D from Median : ( Variables must be arranged in an ascending order )

| S.No. | X | \|dx\| |
|---|---|---|
| 1 | 21 | 7 |
| 2 | 22 | 6 |
| 3 | 24 | 1 |
| 4 | 25 | 7 |
| 5 | 27 | 2 |
| 6 | 29 | 4 |
| 7 | 30 | 1 |
| 8 | 33 | 6 |
| 9 | 34 | 5 |
| 10 | 35 | 3 |

$$N=10 \qquad\qquad\qquad 42$$

$$\text{Median (M )} = \frac{N+1}{2} \text{nd item} = \frac{10 +1}{2} \text{nd item} = 5.5\text{th item}$$

= 5th item + 50% of ( 29-27)

= 27 + 1 = 28

$$M.D = \frac{\in| dx |}{N}$$

$$= \frac{42}{10}$$

= 4.2

$$Coefficient = \frac{M.D}{Average}$$

$$= \frac{4.2}{28} = 0.15$$

**Example 10**

Find out the M.D and its coefficient from Mean and Median from following data.

(x) :  3.2  6.7  4.5  9.4  8.6  6.8  1.3  0.9  4.1  2.0

**Solution :** Computation of Mean Deviation from Mean

| S.No. | X | \|dx\| |
|---|---|---|
| 1 | 3.2 | 1.55 |
| 2 | 6.7 | 1.95 |
| 3 | 4.5 | 0.25 |
| 4 | 9.4 | 4.65 |
| 5 | 8.6 | 3.85 |
| 6 | 6.8 | 2.05 |
| 7 | 1.3 | 3.45 |
| 8 | 0.9 | 3.85 |
| 9 | 4.1 | 0.65 |
| 10 | 2.0 | 2.75 |
| N=10 | 47.5 | 25.00 |

$$a = \frac{x}{N} = \frac{47.5}{10} = 4.75$$

$$\text{M.D} = \frac{\in|dx|}{N} = \frac{25}{10} = 2.5$$

$$\text{Coefficient} = \frac{\text{M.D}}{\text{Average}} = \frac{2.5}{4.75} = 0.526$$

From Median : ( Ascending order )

| S.No. | X | \|dx\| |
|-------|-----|-------|
| 1 | 0.9 | 3.40 |
| 2 | 1.3 | 3.00 |
| 3 | 2.0 | 2.30 |
| 4 | 3.2 | 1.10 |
| 5 | 4.1 | 0.20 |
| 6 | 4.5 | 0.20] |
| 7 | 6.7 | 2.40 |
| 8 | 6.8 | 2.50 |
| 9 | 8.6 | 4.80 |
| 10 | 9.4 | 5.10 |
| N=10 | | 24.50 |

$$\text{Median (M)} = \frac{N+1}{2} \text{ nd item} = \frac{10+1}{2} \text{ nd item} = 5.5\text{th item}$$

= 5th item + 50% of ( 4.5-4.1) = 4.1 + 0.2 = 4.3

$$\text{M.D} = \frac{\in|dx|}{N} = \frac{24.50}{10} = 2.45$$

$$\text{Coefficient} = \frac{\text{M.D}}{\text{Average}} = \frac{2.45}{4.3} = 0.5698$$

**Discrete series - Mean Deviation :**

$$\in|fdx|$$

M.D   = _____N

Where MD = Mean Deviation

$\in$|fdx| = Total of the deviations taken from the average by ignoring the signs (+and –) multiplied with the respective frequencies.
N= Total of the frequency

First of all an average (either mean or Median) must be calculated.

$$\text{Arithmetic Mean (a)} = \frac{\in fdx}{N}$$

$$\text{Median (M)} = \frac{N+1}{2} \text{ nd item}$$

This item must be identified in the 'cf' and corresponding variable must be taken as Median

       Then deviation must be taken from the average by ignoring plus and minus. The deviations must be multiplied with the respective frequencies (fdx). The total of this fdx ($\in$fdx) must be devided with the total of the frequency.

$$\text{Coefficient of M.D.} = \frac{M.D}{Average}$$

**Example 11 :**

Find out the Mean Deviation from Mean & Median from the following data.

| (x) : | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|-------|----|----|----|----|----|----|----|----|----|----|
| (f) : | 4  | 7  | 12 | 13 | 15 | 16 | 14 | 9  | 8  | 2  |

**Solution :** From Arithmetic Mean

| X | f | dx | fdx | \|dx\| | \|fdx\| |
|---|---|-----|------|-------|--------|
| 21 | 4 | -5 | -20 | 4.39 | 17.56 |
| 22 | 7 | -4 | -28 | 3.39 | 23.73 |
| 23 | 12 | -3 | -36 | 2.39 | 28.68 |
| 24 | 13 | -2 | -26 | 1.39 | 18.07 |
| 25 | 15 | -1 | -15 | 0.39 | 5.85 |
| 26 | 16 | 0 | 0 | 0.61 | 9.76 |
| 27 | 14 | +1 | 14 | 1.61 | 22.54 |
| 28 | 9 | +2 | 18 | 2.61 | 23.49 |
| 29 | 8 | +3 | 24 | 3.61 | 28.88 |
| 30 | 2 | +4 | 8 | 4.61 | 9.22 |

|  |  |  |
|---|---|---|
| N=100 | -61 | 187.78 |

$$a = x + \frac{\in fdx}{N} = 26 + \frac{\in 61}{100} = 26 - 0.61 = 25.39$$

$$M.D = \frac{\in |fdx|}{N} = \frac{187.78}{100} = 1.8778$$

$$\text{Coefficient} = \frac{M.D}{\text{Average}} = \frac{1.8778}{25.39} = 0.074$$

**From Median :**

| X | f | Cf | \|dx\| | \|fdx\| |
|---|---|---|---|---|
| 21 | 4 | 4 | 4 | 16 |
| 22 | 7 | 11 | 3 | 21 |
| 23 | 12 | 23 | 2 | 24 |
| 24 | 13 | 36 | 1 | 13 |
| 25 | 15 | 51 | 0 | 0 |
| 26 | 16 | 67 | 1 | 16 |
| 27 | 14 | 81 | 2 | 28 |
| 28 | 9 | 90 | 3 | 27 |
| 29 | 8 | 98 | 4 | 32 |
| 30 | 2 | 100 | 5 | 10 |
| N=100 |  |  |  | 187 |

$$\text{Median position} = \frac{N+1}{2} \text{ nd item} = \frac{100+1}{2} \text{ nd item} = 50.5\text{th item}$$

It lies in the of 'cf' 51 and the corresponding variable is 25

Then the median = 25

$$M.D = \frac{\in |fdx|}{N} = \frac{187}{100} = 1.87$$

$$\text{Coefficient} = \frac{\text{M.D}}{\text{Average}} = \frac{1.87}{25} = 0.074$$

**Continuous series - Mean Deviation :**

$$\text{M.D} = \frac{\in | fdx |}{N}$$

Where MD = Mean Deviation

∈ |fdx| = Total of the deviations taken from the average by ignoring the signs

multiplied with the respective frequency N= Total of the
frequency

First of all an average (Arithermetic mean (a) or Median (M) ) must be calculated. Deviations must be calculated from the average ( by ignoring plus & Minus) to the other variables (|dx|). The deviations must be multiplied with respective frequencies ( |f dx|). The total of this |fdx| must be devided with the total of the frequency (N)

$$\text{Arithmetic Mean (a)} = x + \frac{\in fdx}{N} \times i$$

Median (M) = ＿ nd item, this item must be identified in the 'cf' the corresponding class is to 2 be taken as median class.

$$\text{Then M} = 1 + \frac{c \times i}{f} \text{ is to be applied.}$$

$$\text{Coefficient of M.D.} = \frac{\text{M.D}}{\text{Average}}$$

**Example 12 :**

Find out the Mean Deviation from Mean & Median and its Co-efficient.

| (x) : | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 | 80-90 | 90-100 |
|---|---|---|---|---|---|---|---|---|---|---|
| (f) : | 3 | 6 | 8 | 13 | 16 | 18 | 15 | 12 | 6 | 3 |

**Solution :** From Arithmetic Mean

| X | f | Mv | | dx | fdx | \|dx\| | \|fdx\| |
|---|---|---|---|---|---|---|---|
| 0-10 | 3 | | 5 | -5 | -15 | 46.4 | 139.2 |
| 10-20 | 6 | 15 | | -4 | -24 | 36.4 | 218.4 |
| 20-30 | 8 | 25 | | -3 | -24 | 26.4 | 211.2 |
| 30-40 | 13 | 35 | | -2 | -26 | 16.4 | 213.2 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 40-50 | 16 | 45 | -1 | -16 | 6.4 | 102.4 |
| 50-60 | 18 | 55 | 0 | 0 | 3.6 | 64.8 |
| 60-70 | 15 | 65 | +1 | 15 | 13.6 | 204.0 |
| 70-80 | 12 | 75 | +2 | 24 | 23.6 | 283.2 |
| 80-90 | 6 | 85 | +3 | 18 | 33.6 | 201.6 |
| 90-100 | 3 | 95 | +4 | 12 | 43.6 | 130.8 |
| | N=100 | | | -36 | | 1768.8 |

$$a = x + \frac{fdx}{N} \times i$$

$$= 55 + \frac{36}{100} \times 10$$

$$= 55 - 3.6$$

$$= 51.4$$

$$M.D = \frac{|fdx|}{N}$$

$$= \frac{1768.8}{100} = 17.688$$

$$Coefficient = \frac{M.D}{Average}$$

$$= \frac{17.688}{51.4} = 0.34$$

**From Median :**

| X | f | cf | Mv | \|dx\| | \|fdx\| |
|---|---|---|---|---|---|
| 0-10 | 3 | 3 | 5 | 47.22 | 141.66 |
| 10-20 | 6 | 9 | 15 | 37.22 | 223.32 |
| 20-30 | 8 | 17 | 25 | 27.22 | 217.76 |
| 30-40 | 13 | 30 | 35 | 17.22 | 223.86 |
| 40-50 | 16 | 46 | 45 | 7.22 | 115.52 |

| 50-60 | 18 | 64 | 55 | 2.78 | 50.04 |
| 60-70 | 15 | 79 | 65 | 12.78 | 191.70 |
| 70-80 | 12 | 91 | 75 | 22.78 | 273.36 |
| 80-90 | 6 | 97 | 85 | 32.78 | 196.68 |
| 90-100 | 3 | 100 | 95 | 42.78 | 128.34 |
|  | N=100 |  |  |  | 1762.24 |

$$\text{Median position} = \frac{N}{2} \text{ nd item} = \frac{100}{2} \text{ nd item} = \text{50th item}$$

It lies in the 'cf' 64 and the corresponding class is 50-60

Median class = 50-60

$$= 1 + \frac{c \times i \ M}{f}$$

$$= 50 + \frac{(50 - 46) \times 10}{18}$$

$$= 50 + 2.22$$

$$= 52.22$$

$$M.D = \frac{|fdx|}{N} = \frac{1762.24}{100} = 17.6224$$

$$\text{Coefficient of MD} = \frac{M.D}{\text{Average}} = \frac{17.6224}{52.22} = 0.34$$

**Example 13**

Find out the M.D. and its Coefficient from mean & Median.

| Marks | No. of Students |
|---|---|
| Less than 10 | 3 |
| Less than 20 | 10 |
| Less than 30 | 19 |
| Less than 40 | 32 |
| Less than 50 | 51 |
| Less than 60 | 68 |

| | | |
|---|---|---|
| Less than 70 | 82 | |
| Less than 80 | 94 | |
| Less than 90 | 98 | |
| Less than 100 | 100 | |

**Solution :** From Arithmetic Mean

| X | f | Mv | dx | fdx | \|dx\| | \|fdx\| |
|---|---|---|---|---|---|---|
| 0-10 | 3 | 5 | -4 | -12 | 44.3 | 132.9 |
| 10-20 | 7 | 15 | -3 | -21 | 34.3 | 240.1 |
| 20-30 | 9 | 25 | -2 | -18 | 24.3 | 218.7 |
| 30-40 | 13 | 35 | -1 | -13 | 14.3 | 185.9 |
| 40-50 | 19 | 45 | 0 | 0 | 4.3 | 81.7 |
| 50-60 | 17 | 55 | +1 | +17 | 5.7 | 96.9 |
| 60-70 | 14 | 65 | +2 | +28 | 15.7 | 219.8 |
| 70-80 | 12 | 75 | +3 | +36 | 25.7 | 308.4 |
| 80-90 | 4 | 85 | +4 | +16 | 35.7 | 142.8 |
| 90-100 | 2 | 95 | +5 | +10 | 45.7 | 87.4 |
| | 100 | | | 43 | | 1714.6 |

$$a = x + \frac{fdx}{N} \times i$$

$$= 45 + \frac{43}{100} \times 10$$

$$= 45 + 4.3$$

$$= 49.3$$

$$M.D = \frac{|fdx|}{N}$$

$$= \frac{1714.6}{100} = 17.146$$

$$Coefficient = \frac{M.D}{Average}$$

$$= \frac{17.146}{49.3} = 0.3477$$

**From Median :**

| X | f | cf | Mv | \|dx\| | \|fdx\| |
|---|---|----|----|--------|---------|
| 0-10 | 3 | 3 | 5 | 44.47 | 133.41 |
| 10-20 | 7 | 10 | 15 | 34.47 | 241.29 |
| 20-30 | 9 | 19 | 25 | 24.47 | 220.23 |
| 30-40 | 13 | 32 | 35 | 14.47 | 188.11 |
| 40-50 | 19 | 51 | 45 | 4.47 | 84.93 |
| 50-60 | 17 | 68 | 55 | 5.53 | 94.01 |
| 60-70 | 14 | 82 | 65 | 15.53 | 217.42 |
| 70-80 | 12 | 94 | 75 | 25.53 | 306.36 |
| 80-90 | 4 | 98 | 85 | 35.53 | 142.12 |
| 90-100 | 2 | 100 | 95 | 45.53 | 91.06 |
| | N=100 | | | | 1718.94 |

$$\text{Median position} = \frac{N}{2} \text{ nd item} = \frac{100}{2} \text{ nd item} = 50\text{th item}$$

It lies in the 'cf' 51 and the corresponding class is 40-50

Median class = 40-50

$$M = l + \frac{c \times i}{f}$$

$$= 40 + \frac{(50 - 32) \times 10}{19}$$

$$= 40 + \frac{186}{19}$$

$$= 40 + 9.473$$

$$= 49.473$$

$$M.D = \frac{|fdx|}{N} = \frac{1718.94}{100} = 17.1894$$

$$\text{M.D} \quad 17.1894$$
$$\text{Coefficient} = \text{Average} = 49.473 = 0.3474$$

## 10. 12 MERITS OF MEAN DEVIATION :

1. It is rigidly defined easy to compute and understand.

2. It takes all the items into consideration and gives weight to deviation according to their size

3. It is less affected by extreme values of variables

4. It removes all the irregularities by obtaining deviation and provides a correct measure.

## 10.13 DEMERITS OF MEAN DEVIATION :

1. It does not lend itself readily to algebraic treatment

2. It ignores the negative deviation and treats them as positive which is not justified mathematically

3. It is rarely used in social sciences

4. It is not suitable when the class intervals are open end.

## 10.14 SUMMARY :

Range is the difference between the Highest value and least value. But it is not a stable measure and has many limitations such as fluctuations of sampling . Quartile Deviation is better than Range. But here all the items are not taken into account. It also suffers from sampling instability. Mean Deviation is better than quartile Deviation. But it is also not capable of further Algebraic treatment, although it takes into account all the terms but still if the extreme values are big, it will desort the result. More over it ignores + signs.

## 10.15 QUESTIONS :

1. What is meant by measures of dispersion ?

2. What are the differences between the Measures of central tendency and Measures of Dispersion ?

3. What are the objectives of measuring the second order average ?

4. Define variation or Dispersion or scatterdness

5. Name various methods of measuring dispersion.

6. Define Range . Is it positional measure ? How ?

7. What is coeffcient of Range ? Narrate the formula ?

8. What are the merits and demerits of Range ?

9. Define Semi - inter Quartile Range.

10. What are the objectives of computing the Quartile Deviation ?

11. What are the merits and demerits of Quartile deviation ?

12. Define Mean Deviation or Average Deviation

13. What is meant by Coefficient of Mean Deviation ?

14. Explain the method of calculation of M.D from Mean

15. Mean Deviation is free from all the short comings or Range Q.D. and hence is a super measure of variation. Discuss.

16. What are the merits and demerits of Mean Devaition ?

## 10.16  EXERCISES :

1. Find out of the Range and coefficient of Range from the following data.

   wages ( Rs) = 27, 31, 32, 28, 40,39,37,34,30,29

2. The earnings of a worker in a week were as under. Find out Range & its co efficient .

   Earnings ( Rs ) = 26,35,41,45,32,29.

3. Find the range and coefficient of Range for following data.

   | (x) : | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 |
   |---|---|---|---|---|---|---|---|---|
   | (f) : | 4 | 7 | 21 | 47 | 53 | 24 | 12 | 6 |

4. Calculate Range and its coefficient from the following data

   | (x) : | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 |
   |---|---|---|---|---|---|---|
   | (f) : | 4 | 9 | 16 | 21 | 13 | 6 |

5. Compute Quartile Deviation and its coefficient from the following data.Marks of 11 students : 75, 43, 86, 21, 12, 3, 35, 57, 67, 94, 60

6. Calculate Quartile Deviation and its Coefficient from the data given below : Wages 12 workess : 91, 98, 99 , 90, 89, 94, 93, 97, 96, 94, 95, 92

7. From the following data compute Q.D. and its coefficient

   | (x) : | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 |
   |---|---|---|---|---|---|---|---|---|---|---|
   | (f) : | 8 | 16 | 26 | 32 | 36 | 28 | 22 | 18 | 10 | 4 |

8. From the following data compute Q.D. and its coefficient

   | (x) : | 0-5 | 5-10 | 10-15 | 15-20 | 20-25 | 25-30 | 30-35 | 35-40 | 40-45 | 45-50 |
   |---|---|---|---|---|---|---|---|---|---|---|
   | (f) : | 6 | 9 | 13 | 18 | 21 | 19 | 17 | 16 | 12 | 8 |

9. Find out Q.D and its coefficient from the following data

   | (x) : | 0-2 | 2-4 | 4-6 | 6-8 | 8-10 | 10-12 | 12-14 | 14-16 | 16-18 | 18-20 |
   |---|---|---|---|---|---|---|---|---|---|---|
   | (f) : | 7 | 13 | 18 | 23 | 22 | 20 | 16 | 15 | 14 | 11 |

10. Find out Q.D and its coefficient from the following data

    | (x) : | 1-10 | 11-20 | 21-30 | 31-40 | 41-50 | 51-60 | 61-70 | 71-80 | 81-90 | 91-100 |
    |---|---|---|---|---|---|---|---|---|---|---|
    | (f) : | 5 | 9 | 13 | 19 | 22 | 23 | 21 | 18 | 17 | 12 |

11. Find Q.D and its coefficient

Mid values of classes -

| 115 | 125 | 135 | 145 | 155 | 165 | 175 | 185 | 195 | 205 |

Frequencey -

| 6 | 12 | 13 | 21 | 25 | 24 | 23 | 21 | 19 | 7 |

12. Find Q.D and its coefficient

Mid values of classes -

| 2.5 | 7.5 | 12.5 | 17.5 | 22.5 | 27.5 | 32.5 | 37.5 | 42.5 | 47.5 |

Frequencey -

| 6 | 9 | 13 | 18 | 17 | 16 | 12 | 8 | 6 | 5 |

13. Find out Q.D and its coefficient

| Marks | No. of Students |
|-------|-----------------|
| Less than 10 | 4 |
| Less than 20 | 11 |
| Less than 30 | 24 |
| Less than 40 | 41 |
| Less than 50 | 57 |
| Less than 60 | 71 |
| Less than 70 | 83 |
| Less than 80 | 92 |
| Less than 90 | 97 |
| Less than 100 | 100 |

14. Find out Mean Deviation and it's coefficient from Mean and Median.

| x - | 75 | 64 | 79 | 67 | 70 | 61 | 68 | 82 | 63 | 71 |

15. Find out Mean Deviation at it's coefficient from Mean and Median.

x   -12.6, 13.9,19.8, 14.7, 11.5,17.3, 16.2,10.1, 15.4,18.0, 16.7

16. Find Q.D and its coefficient

| x - | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 |
| f - | 3 | 8 | 13 | 19 | 16 | 14 | 12 | 9 | 4 | 2 |

17. Find out mean Deviation from Median.

| x - | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| f - | 5 | 8 | 13 | 16 | 18 | 14 | 12 | 8 | 4 | 2 |

18. Find out Mean Deviation from Mean & Median Also find out the coefficent.

x   : 0-5    5-10    10-15    15-20    20-25    25-30    30-35    35-40    40-45    45-50

   (f) :     13   18     19    23    26     24      22      21     19    15

19. Find out Mean Deviation and its coefficient from Mean & Median.

x     : 0-9   10-19   20-29   30-39   40-49   50-59   60-69   70-79   80-89   90-99

   (f) :     5    9     12    13    14     16     13      9      7     2

20. Find out Mean Deviation and its coefficeint from the following data.

| Marks | No. of Students |
|---|---|
| Above 0 | 100 |
| Above 10 | 97 |
| Above 20 | 89 |
| Above 30 | 76 |
| Above 40 | 60 |
| Above 50 | 42 |
| Above 60 | 28 |
| Above 70 | 16 |
| Above 80 | 7 |
| Above 90 | 3 |

21. Find out Mean Deviation and its coefficient

  (x) :     0-99   100-199     200-299        300-399        400-499

  (f) :     2      8        13          15          18

  (x) : 500-599 600-699     700-799     800-899     900-999

  (f) :     14    11       9        7         3

**T. Nageswara Rao**

**Vice - Principal**

**Hindu College, Guntur.**

# LESSON - 11

# MEASURES OF DISPERSION -STANDARD DEVIATION

**OBJECTIVES :**

By the study of this lesson, you will be able to understand the meaning and methods of computation of standard deviation, Coefficient of variation in individual , Discrete & Continuous series with examples.

**STRUCTURE OF THE LESSON :**

11.1    Introduction

11.2    Computation of Standard Deviation

11.3    Co efficient of Variation

11.4    Examples

11.5    Merits of standard Deviation

11.6    Demerits of standard Deviation

11.7    Summary

11.8    Questions

11.9    Exercises

## 11.1    INTRODUCTION :

" Standard deviation " is the root of the sum of the squares of the deviations divided by their number. It is also called Mean " Error Deviation " " Mean square Error Deviation " or " Root MeanSquare Deviation ". It is Second moment of a dispersion. Since the sum of the squares of the deviations from the Mean is minimum, the deviations are taken only from mean ( but not from median or mode ).

Standard Deviation is the root - mean - square average of all the deviations from the mean. It is proposed by " Prof - karl pearson " in 1893 and it is denoted by '$\sigma$ '( Sigma )

## 11.2    COMPUTATION OF STANDARD DEVIATION :

Individual Series **:**

$$\sigma = \sqrt{\frac{\in dx^2}{N} - \left|\left(\frac{\in dx}{N}\right)^2\right|}$$

Where

$\sigma$  = Standard Deviation

$\in dx$ = Total of the deviations taken from the assumed mean

$\in dx^2$ = Total of the squares of the deviations taken from the assumed meanN = Number of variables.

## 11.3    CO - EFFICIENT OF VARIATION :

It is the relative measure of dispersion in which the variation is expressed in percentage. It is often used to have the comparative study of the dispersion of two or more series in the same or different units. It is the percentage variation in the mean, where as the standard deviation is the total variation in the mean. This relative measure of dispersion implies the ratio of standard deviation to the mean signifying the percentage.

Co efficient of variation = $\frac{\sigma}{a} \times 100$ Where

$\sigma$ = Standard deviation

It is helpful in knowing the consistency of items of the series. If the value so arrived is greater ( more thean 50%) the result signifies the lower degree of consistency. If the value so arrived is smaller ( Less than 50% ) the result signifies upper degree of consistency.

## 11.4    EXAMPLES :

**Example 1 :** Find out standard Deviation and coefficient of variation

x :      24    31    27    25    28    20    29    23    22    30

**Solution :**

| X | dx | dx² |
|---|---|---|
| 24 | -3 | 9 |
| 31 | +4 | 16 |
| 27 | 0 | 0 |
| 25 | -2 | 4 |
| 28 | +1 | 1 |
| 20 | -7 | 49 |
| 29 | +2 | 4 |
| 23 | -4 | 16 |
| 22 | -5 | 25 |
| 30 | +3 | 9 |
| N=10 | -11 | 133 |

$$\sigma = \sqrt{\frac{\in dx^2}{N} - \left(\frac{\in dx}{N}\right)^2}$$

$$= \sqrt{\frac{133}{10} - \left(\frac{-11}{10}\right)^2}$$

$$= \sqrt{12.09}$$

$$= \quad \cancel{13.3 = 1.21}$$

$$a = x + \frac{\in dx}{N}$$

$$= 27 + \frac{-11}{10} = 27 - 1.1 = 25.9$$

Coefficient of variation $= \sigma$

$$= \frac{3.476}{25.9} \times 100 = 13.42 \%$$

**Example 2 :** The prices shares of 2 companies were as under. Which is more variable ?

Company A : 12　15　21　16　9　13　10　17　14 21　11 8

Company B : 107　109　100　111　97　93　96　104　101 108　106

**Solution :**

| X | dx | dx² | y | dy | dy² |
|---|---|---|---|---|---|
| 12 | -3 | 9 | 107 | +7 | 49 |
| 15 | 0 | 0 | 109 | +9 | 81 |
| 21 | +6 | 36 | 100 | 0 | 0 |
| 16 | +1 | 1 | 111 | +11 | 121 |
| 9 | -6 | 36 | 97 | -3 | 9 |
| 13 | -2 | 4 | 93 | -7 | 49 |
| 10 | -5 | 25 | 96 | -4 | 16 |
| 17 | +2 | 4 | 104 | +4 | 16 |
| 14 | -1 | 1 | 101 | +1 | 1 |
| 21 | +6 | 36 | 108 | +8 | 64 |
| 11 | -4 | 16 | 106 | +6 | 36 |
| 8 | -7 | 49 | 105 | +5 | 25 |
| N=12 | -13 | 217 | N12 | 37 | 407 |

**X - Series** :

$$\sigma = \sqrt{\frac{\in dx^2}{N} - \left| \frac{\in dx}{N} \right|^2}$$

$$= \sqrt{\frac{217}{12} - \left(\frac{-13}{12}\right)^2}$$

$$= \sqrt{18.08 - (-1.08)^2}$$

$$= \sqrt{18.08 - 1.1664} \quad \sqrt{16.9136} \quad = 4.1126$$

$$a = x + \frac{\in dx}{N}$$

$$= 15 + \frac{-13}{12} = 15 - 1.08 = 13.92$$

Coefficient of variation $= \dfrac{\sigma}{a} \times 100$

$$= \frac{4.1126}{13.92} \times 100 = 29.54\%$$

**Y - Series** :

$$\sigma = \sqrt{\frac{\in dy^2}{N} - \left(\frac{\in dy}{N}\right)^2}$$

$$= \sqrt{\frac{467}{12} - \left(\frac{-37}{12}\right)^2} \quad \sqrt{38.917 - (3.083)^2}$$

$$= \sqrt{38.917 - 9.505}$$

$$= \sqrt{29.412} = 5.423$$

$$a = y + \frac{\in dy}{N}$$

$$= 100 + \frac{37}{12} = 100 + 3.083$$

$$= 103.083$$

Coefficient of variation $= \dfrac{\sigma}{a} \times 100$

$$= \qquad \frac{5.423}{103.083} \qquad \text{x100} \qquad = 5.2608\ \%$$

The coefficient of variation is more in company A. There fore it can be said that the prices ofshares of company A are more variable.

**Example 3 :** The marks secured by two students A&B in 10 examinations were as under. Find out who is more clever ? ( or Find out who is more consistent )

Mark A :        42    70    36    30    48    45    34    50    60    25

Mark B :        55    95    42    20    60    50    48    70    80    10

**Solution :**

| X | dx | dx² | y | dy | dy² |
|---|---|---|---|---|---|
| 42 | -3 | 9 | 55 | +5 | 25 |
| 70 | +25 | 625 | 95 | +45 | 2025 |
| 36 | -9 | 81 | 42 | -8 | 64 |
| 30 | -15 | 225 | 20 | -30 | 900 |
| 48 | +3 | 9 | 60 | +10 | 100 |
| 45 | 0 | 0 | 50 | 0 | 0 |
| 34 | -11 | 121 | 48 | -2 | 4 |
| 50 | +5 | 25 | 70 | +20 | 400 |
| 60 | +15 | 225 | 80 | +30 | 900 |
| 25 | -20 | 400 | 10 | -40 | 1600 |
| N=10 | -10 | 1720 | N=10 | 30 | 6018 |

**X - Series :**

$$\sigma = \sqrt{\frac{\in dx^2}{N} - \left| \frac{\in dx}{N} \right|^2}$$

$$= \sqrt{\frac{1720}{10} - \left| \frac{-10}{10} \right|^2}$$

$$= \sqrt{172 - 1} = \sqrt{171} = 13.08$$

$$a = x + \frac{\in dx}{N}$$

$$= 45 + \frac{-10}{} \, 10 \qquad = 45 - 1 \; = 44$$

Coefficient of variation =

$$\frac{\sigma}{a} \times 100$$

$$= \frac{13.08}{44} \times 100$$

$= 29.727 \, \%$

**Y - Series** :

$$\sigma = \sqrt{\frac{\in dy^2}{N} - \left| \left( \frac{\in dy}{N} \right) \right|^2}$$

$$= \sqrt{\frac{6018}{10} - \left| \left( \frac{30}{10} \right) \right|^2}$$

$$= \sqrt{601.8 - 9}$$

$$= \sqrt{592.8}$$

$= 24.347$

$$a = y + \frac{\in dy}{N}$$

$$= 50 + \frac{30}{10} = 50 + 3$$

$$= 53$$

Coefficient of variation = $\frac{\sigma}{a} \times 100$

$$= \frac{24.347}{53} \times 100$$

$= 45.94$

The coefficient of variation in series x is smaller than Series Y. Therefore it can be concludedthat Mr A ( x series ) is cleverer than Mr. B ( Y series ) or A is more consistent.

**Discrete Series :**

$$\sigma = \sqrt{\frac{\in fdx^2}{N} - \left|\left(\frac{\in fdx}{N}\right)\right|^2}$$

Where  x i

$\sigma$ = Standard Deviation

$\in fdx$ = Total of the deviations taken from the assumed mean, multiplied with the respective frequencies.

$\in fdx^2$ = Total of the squares of the deviations taken from the assumed mean, multiplied with the respective frequencies.

N = Number of variables.

i = Interval ( common factor )

Coefficient of variation = $\dfrac{\sigma}{a}$ x100

$$a = x + \frac{\in fdx}{N} x i$$

**Example 4 :** Compute standard Deviation and its coefficient of variation from the following data.

| x: 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|------|---|---|----|----|----|----|----|---|---|----|
| f : 3 | 5 | 8 | 11 | 13 | 16 | 14 | 12 | 9 | 6 | 3 |

**Solution :**

| X | f | dx | fdx | fdx² |
|----|------|----|-----|------|
| 0 | 3 | -5 | -15 | 75 |
| 1 | 5 | -4 | -20 | 80 |
| 2 | 8 | -3 | -24 | 72 |
| 3 | 11 | -2 | -22 | 44 |
| 4 | 13 | -1 | -13 | 13 |
| 5 | 16 | 0 | 0 | 0 |
| 6 | 14 | +1 | 14 | 14 |
| 7 | 12 | +2 | 24 | 48 |
| 8 | 9 | +3 | 27 | 81 |
| 9 | 6 | +4 | 24 | 96 |
| 10 | 3 | +5 | 15 | 75 |
| | N=100 | | 10 | 598 |

$$\sigma = \sqrt{\frac{\in fdx^2}{N} - \left(\frac{\in fdx}{N}\right)^2}$$

$$= \sqrt{\frac{598}{100} - \left(\frac{10}{100}\right)^2}$$

$$= \sqrt{5.98 - 0.01}$$

$$= \sqrt{5.97} = 2.443$$

$$a = x + \frac{\in fdx}{N}$$

$$= \frac{5 + 10}{100}$$

$$= 5.1$$

Coefficient of variation = $\dfrac{\sigma}{a} \times 100$

$$= \frac{2.443}{5.1} \times 100$$

$$= 47.90\%$$

**Example 5 :** Compute standard Deviation and its coefficient of variation

| x: | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 |
|----|---|----|----|----|----|----|----|----|----|----|
| f : | 3 | 7 | 11 | 13 | 17 | 16 | 12 | 9 | 8 | 4 |

**Solution :**

| X | f | dx | fdx | fdx$^2$ |
|---|---|----|-----|---------|
| 5 | 3 | -4 | -12 | 48 |
| 10 | 7 | -3 | -21 | 63 |
| 15 | 11 | -2 | -22 | 44 |
| 20 | 13 | -1 | -13 | 13 |
| 25 | 17 | 0 | 0 | 0 |
| 30 | 16 | +1 | 16 | 16 |
| 35 | 12 | +2 | 24 | 48 |
| 40 | 9 | +3 | 27 | 81 |
| 45 | 8 | +4 | 32 | 128 |
| 50 | 4 | +5 | 20 | 100 |
| | N=100 | | 51 | 541 |

$$\sigma = \sqrt{\frac{\in fdx^2}{N} - \left|\left(\frac{\in fdx}{N}\right)\right|^2} \quad x\, i$$

$$= \sqrt{\frac{541}{100} - \left|\left(\frac{51}{100}\right)\right|^2} \quad x\, i$$

$$= \sqrt{5.41 - (0.51)^2} \quad x\, i$$

$$= \sqrt{5.41 - 0.2601} \quad x\, i$$

$$= \sqrt{\quad\quad}$$

$$= 11.35 \ x\, i = 2.27 \ x \ 5$$

$$a = x + \frac{\in fdx}{N} \ x\, i$$

$$= \frac{25 + 51}{100} \ x \ 5$$

$$= 25 + \frac{255}{100}$$

$$= 25 + 2.55$$

$$= 27.55$$

Coefficient of variation $= \dfrac{}{a}$ x100

$$= \frac{11.35}{27.55} \text{ x100}$$

$$= 41.2\%$$

**Continuous Series :**

$$\sigma = \sqrt{\frac{\in fdx^2}{N} - \left|\frac{\in fdx}{N}\right|^2}$$

Where  x i

$\sigma$ = Standard Deviation

$\in fdx$ = Total of the deviations taken from the assumed mean, multiplied with the respective frequencies.

$\in fdx^2$ = Total of the squares of the deviations taken from the assumed mean, multiplied with the respective frequencies.

N = Number of variables.

i = Interval ( common factor )

**Note :** In continuous series the classes must be converted into Mid values, assumed mean shall be taken from the Mid values and the deviation shall be taken from theassumed mean to the other Mid values.

Coefficient of variation $= \dfrac{\sigma}{a}$ x100

$$a = x + \frac{\in fdx}{N} \text{ x i}$$

**Example 6 :** Compute standard Deviation and its coefficient

| (x) : | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 | 80-90 | 90-100 |
|-------|------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| (f) : | 3 | 6 | 8 | 13 | 16 | 15 | 14 | 12 | 9 | 4 |

**Solution :** From Arithmetic Mean

| X | f | Mv | dx | fdx | fdx$^2$ |
|---|---|---|---|---|---|
| 0-10 | 3 | 5 | -4 | -12 | 48 |
| 10-20 | 6 | 15 | -3 | -18 | 54 |
| 20-30 | 8 | 25 | -2 | -16 | 32 |
| 30-40 | 13 | 35 | -1 | -13 | 13 |
| 40-50 | 16 | 45 | 0 | 0 | 0 |
| 50-60 | 15 | 55 | +1 | +15 | 15 |
| 60-70 | 14 | 65 | +2 | +28 | 56 |
| 70-80 | 12 | 75 | +3 | +36 | 108 |
| 80-90 | 9 | 85 | +4 | +36 | 144 |
| 90-100 | 4 | 95 | +5 | +20 | 100 |
| | N=100 | | | 76 | 570 |

**X - Series**
:

$$\sigma = \sqrt{\frac{\in fdx^2}{N} - \left|\left(\frac{\in fdx}{N}\right)\right|^2} \quad x \ i$$

$$= \sqrt{\frac{570}{100} - \left|\left(\frac{76}{100}\right)\right|^2} \quad x \ i$$

$$= \sqrt{5.7 - 0.5776} \quad x \ i$$

$$= \sqrt{5.1224} \quad x \ i$$

$$= 2.2632 \ x \ 10 \ = 22.632$$

$$a = \frac{\sigma}{a} \ x100$$

$$= \frac{22.632}{52.6} \ x100$$

$$= 43.03\%$$

**Example 7 :** Find out the standard Deviation and its coefficient of variation.

| Marks | No. of Students |
|---|---|
| More than 0 | 100 |
| More than 10 | 97 |
| More than 20 | 89 |
| More than 30 | 77 |
| More than 40 | 64 |
| More than 50 | 57 |
| More than 60 | 42 |
| More than 70 | 28 |
| More than 80 | 17 |
| More than 90 | 5 |

**Solution :** From Arithmetic Mean

| X | f | Mv | dx | fdx | fdx² |
|---|---|---|---|---|---|
| 0-10 | 3 | 5 | -5 | -15 | 75 |
| 10-20 | 8 | 15 | -4 | -32 | 128 |
| 20-30 | 12 | 25 | -3 | -36 | 108 |
| 30-40 | 13 | 35 | -2 | -26 | 52 |
| 40-50 | 7 | 45 | -1 | -7 | 7 |
| 50-60 | 15 | 55 | 0 | 0 | 0 |
| 60-70 | 14 | 65 | +1 | +14 | 14 |
| 70-80 | 11 | 75 | +2 | +22 | 44 |
| 80-90 | 12 | 85 | +3 | +36 | 108 |
| 90-100 | 5 | 95 | +4 | +20 | 80 |
| | N=100 | | | -24 | 616 |

$$\sigma = \sqrt{\frac{\epsilon fdx^2}{N} - \left|\frac{\epsilon fdx}{N}\right|^2} \times i$$

$$= \sqrt{\frac{616}{100} - \left(\frac{-24}{100}\right)^2} \quad \text{x i}$$

$$= \sqrt{6.16 - (-0.24)^2} \quad \text{x i}$$

$$= \sqrt{6.16 - 0.0576} \quad \text{x i}$$

$$= \sqrt{6.1024} \quad \text{x i}$$

$$= 2.47 \text{ x } 10 = 24.7$$

$$a = x + \frac{\in fdx}{N} \quad \text{x i}$$

$$= 55 + \frac{-24}{100} \text{x } 10$$

$$= 55 - 2.4 = 52.6$$

$$\text{Coefficient of variation} = \frac{\sigma}{a} \text{x} 100$$

$$= \frac{24.7}{52.6} \text{x} 100$$

$$= 46.96\%$$

**Example 8 :** Which of the following two series is more consistent in value ?

| x : | 0-5 | 5-10 | 10-15 | 15-20 | 20-25 | 25-30 | 30-35 | 35-40 | 40-45 | 45-50 |
|-----|-----|------|-------|-------|-------|-------|-------|-------|-------|-------|
| $f_1$ : | 13 | 17 | 19 | 23 | 27 | 25 | 22 | 21 | 19 | 14 |
| $f_2$ : | 15 | 16 | 21 | 24 | 28 | 25 | 23 | 18 | 16 | 14 |

**Solution :**

**$f_1$**

| X | f | Mv | dx | fdx | fdx² |
|------|------|------|------|------|------|
| 0-5 | 13 | 2.5 | -4 | -52 | 208 |
| 5-10 | 17 | 7.5 | -3 | -51 | 153 |
| 10-15 | 19 | 12.5 | -2 | -38 | 76 |
| 15-20 | 23 | 17.5 | -1 | -23 | 23 |
| 20-25 | 27 | 22.5 | 0 | 0 | 0 |
| 25-30 | 25 | 27.5 | +1 | +25 | 25 |
| 30-35 | 22 | 32.5 | +2 | +44 | 88 |
| 35-40 | 21 | 37.5 | +3 | +63 | 189 |
| 40-45 | 19 | 42.5 | +4 | +76 | 304 |
| 45-50 | 14 | 47.5 | +5 | +70 | 350 |
| | N=200 | | | 114 | 1416 |

**$f_2$**

| X | f | Mv | dx | fdx | fdx² |
|------|------|------|------|------|------|
| 0-5 | 15 | 2.5 | -4 | -60 | 240 |
| 5-10 | 16 | 7.5 | -3 | -48 | 144 |
| 10-15 | 21 | 12.5 | -2 | -42 | 84 |
| 15-20 | 24 | 17.5 | -1 | -24 | 24 |
| 20-25 | 28 | 22.5 | 0 | 0 | 0 |
| 25-30 | 25 | 27.5 | +1 | +25 | 25 |
| 30-35 | 23 | 32.5 | +2 | +46 | 92 |
| 35-40 | 18 | 37.5 | +3 | +54 | 162 |
| 40-45 | 16 | 42.5 | +4 | +64 | 256 |
| 45-50 | 14 | 47.5 | +5 | +70 | 350 |
| | N=200 | | | 85 | 1377 |

**f₁- Series** :

$$\sigma = \sqrt{\frac{\in fdx^2}{N} - \left(\frac{\in fdx}{N}\right)^2} \qquad x\ i$$

$$= \sqrt{\frac{1416}{200} - \left(\frac{114}{200}\right)^2} \qquad x\ i$$

$$= \sqrt{7.08 - 0.3249} \qquad x\ i$$

$$= \sqrt{6.7751}\ x\ i$$

$$= 2.599\ x\ 5 = 12.995$$

$$a = x + \frac{\in fdx}{N}\ x\ i$$

$$= 22.5 + \frac{114}{200} x\ 5$$

$$= 22.5 + \frac{570}{200} = 22.5 + 2.85 = 25.35$$

Coefficient of variation $= \dfrac{\sigma}{a} x 100$

$$= \frac{12.995}{25.35} x 100$$

$$= 51.26\ \%$$

**f₂- Series** :

$$\sigma = \sqrt{\frac{\in fdx^2}{N} - \left(\frac{\in fdx}{N}\right)^2} \qquad x\ i$$

$$= \sqrt{\frac{1377}{200} - \left(\frac{85}{200}\right)^2} \qquad x\ i$$

$$= \sqrt{6.885 - 0.1806} \qquad x\ i$$

$$= \sqrt{6.7644} \quad \times \ i$$

$$= 2.5892 \times 5 = 12.946$$

$$a = x + \frac{\in fdx}{N} \times i$$

$$= 22.5 + \frac{85}{200} \times 5$$

$$= 22.5 + \frac{425}{200} = 22.5 + 2.125 = 24.625$$

$$\text{Coefficient of variation} = \frac{\sigma}{a} \times 100$$

$$= \frac{12.946}{24.625} \times 100 \qquad = 52.57 \ \%$$

The coefficient of variation is small in $f_1$ series. Therefore, it can be said that the series $f_1$ are more consistant in value.

**Example 9 :** The profits and losses of 100 companines in an industry were as under.

Find out the standard Deviation and its coefficient of variation.

| Profits & Losses | No. of companies |
|---|---|
| 4000-5000 | 5 |
| 3000-4000 | 9 |
| 2000-3000 | 12 |
| 1000-2000 | 13 |
| 0-1000 | 17 |
| -1000 - 0 | 16 |
| -2000- -1000 | 14 |
| -3000- -2000 | 8 |
| -4000- -3000 | 4 |
| -5000- -4000 | 2 |
| | N=100 |

**Solution :**

| X | f | Mv | dx | fdx | fdx² |
|---|---|----|----|-----|------|
| 4000-5000 | 5 | 4500 | +4 | +20 | 80 |
| 3000-4000 | 9 | 3500 | +3 | +27 | 81 |
| 2000-3000 | 12 | 2500 | +2 | +24 | 48 |
| 1000-2000 | 13 | 1500 | +1 | +13 | 13 |
| 0-1000 | 17 | 500 | 0 | 0 | 0 |
| -1000 - 0 | 16 | -500 | -1 | -16 | 16 |
| -2000- -1000 | 14 | -1500 | -2 | -28 | 56 |
| -3000- -2000 | 8 | -2500 | -3 | -24 | 72 |
| -4000- -3000 | 4 | -3500 | -4 | -16 | 64 |
| -5000- -4000 | 2 | -4500 | -5 | -10 | 50 |
| | N=100 | | | -10 | 480 |

$$\sigma = \sqrt{\frac{\in fdx^2}{N} - \left|\left(\frac{\in fdx}{N}\right)\right|^2} \quad \times \ i$$

$$= \sqrt{\frac{480}{100} - \left|\left(\frac{-10}{100}\right)\right|^2} \quad \times \ i$$

$$= \sqrt{4.8 - 0.01} \times i$$

$$= \sqrt{4.79} \quad \times \ i$$

$$= 2.1886061 \times 1000 = 2188.61/-$$

$$a = x + \frac{\in fdx}{N} \times i$$

$$= 500 + \frac{-10}{100} \times 1000$$

$$= 500 - 100 = 400/-$$

$$\text{Coefficient of variation} = \frac{\sigma}{a} \times 100$$

$$= \frac{2188.61}{400} \times 100$$

$$= 547.15\%$$

## 11.5 MERITS OF STANDARD DEVIATION :

a) It is based on all the observations given

b) It can be smoothly handled algebraically

c) It is a well defined and definite measure of dispersion.

d) It is of great importance when the comparison is made between variability of twoitem.

## 11.6 DEMERITS OF STANDARD DEVIATION :

1. It is difficult to calculate and understand

2. It gives more weight to extreme values as the deviations are squared.

3. It is not useful in economic studies.

## 11.7 SUMMARY :

Standard Deviation and Coefficient of variation possess all those properties, which a good measure of disperssion should possess. The process of squaring the deviations eliminates the negative signs and thus makes the mathematical manipulation of figures earh.

## 11.8 QUESTIONS :

1. Define standard Deviation ?

2. What is meant by Standard Deviation ?

3. What is meant by Coefficient of Variation ?

4. What are the merits and Demerits of Standard Deviation ?

5. Why the Standard Deviation is better than the other measures of dispersion ?

## 11.9 EXERCISES :

1. Find out Standard Deviation and its coefficient of

variation . (x) : 24    27    23    30    25    29    21    27    26    22

2. Find out Standard Deviation and its coefficient of

variation . (x) : 57    58    52    56    60    55    51    54    53    59

3.          Compute Standard Deviation and its coefficient of

variation . (x) :          85     94     93     90     96     99     98     91     87     86

4.          Calculate Standard Deviation  and its coefficient of variation .

(x) : 345, 352,  341,  350,  355,  357,  354,  344,  348,  349,   341, 346,

5.          The following 2 series were given to you : Which is more consistent in

value. ?(x) : 75        49     56     64     70     65     67     73     58     74

(f) :  160                 170   173   164   167   161   175   177   172 169

6.      The runs scored by two batsmen in 10 one day matches were as under. Who is more consistent ?

          a :       3     87     64     1     12     76     0     50     60     85 96     24

          b :      75     86     63     47     55     60     49     21     13     70 49     2

7.      The prices of the shares of two companies during last 12 months were as under. Which share is more variable in price.

          Company x : 5        12     17     14     13     10     6     8     13     15   16        9

          Company y :45        54     49     50     51     53     42     47     48     55 54        52

8.          Compute Standard Deviation  and its coefficient of variation

          x :       41     42     43     44     45     46     47     48     49     50

          f :        5     9     13     17     18     16     14     7     8     3

9.          Compute Standard Deviation  and its coefficient

          x :        2     4     6     8     10     12     14     16     18     20

          f :        3     7     11     13     17     16     12     9     8     4

10.          Compute Standard Deviation  and its coefficient

(x) :      0-5 5-10 10-15   15-20   20-25     25-30     30-35   35-40   40-45   45-50

(f) :      4        7        8        12        18        17        13        11        64

11.          Find out Standard Deviation  and its coefficient

(x) :20-24   24-28   28-32   32-36 36-40     40-44     44-48   48-52     52-56

(f) :8     12        13        15        18        14     9        7        4

12.          Compute Standard Deviation  and its coefficient of variation

(x) : 0-9  10-19  20-29  30-39    40-49  50-59  60-69  70-79  80-89  90-99

(f) :                    4        8      12      13      15      14      12      9      8
                    5

13.          Compute Standard Deviation  and its coefficient of variation

| (x) : | 0-99 | 100-199 | 200-299 | 300-399 | 400-499 |
|-------|------|---------|---------|---------|---------|
| (f) : | 5 | 7 | 12 | 13 | 16 |
| (x) : | 500-599 | 600-699 | 700-799 | 800-899 | 900-999 |
| (f) : | 17 | 14 | 9 | 5 | 2 |

14.          Find out Standard Deviation  and its coefficient

(x) : 0-4                    5-9  10-14  15-19    20-24    25-29    30-34  35-39  40-44    45-49

(f) :              13      17    19      23      27      25      22      21      19      14

15.          Compute Standard Deviation  and its coefficient of variation

| Marks | No. of Students |
|-------|-----------------|
| Less than 10 | 4 |
| Less than 20 | 11 |
| Less than 30 | 20 |
| Less than 40 | 33 |
| Less than 50 | 50 |
| Less than 60 | 66 |
| Less than 70 | 81 |
| Less than 80 | 93 |
| Less than 90 | 98 |
| Less than 100 | 100 |

16.        Calculate Standard Deviation and its coefficient of variation

| Marks | No. of Students |
|---|---|
| Less than 100 | 100 |
| Less than 90 | 97 |
| Less than 80 | 90 |
| Less than 70 | 77 |
| Less than 60 | 60 |
| Less than 50 | 44 |
| Less than 40 | 30 |
| Less than 30 | 18 |
| Less than 20 | 9 |
| Less than 10 | 4 |

17.        The profits and losses of 100 companies in an industry were as under. Find out thestandard Deviation and its coefficient of variation.

| Profits & Losses | No. of Companies |
|---|---|
| 4000-5000 | 2 |
| 3000-4000 | 7 |
| 2000-3000 | 9 |
| 1000-2000 | 13 |
| 0-1000 | 16 |
| -1000 - 0 | 17 |
| -2000- -1000 | 14 |
| -3000- -2000 | 12 |
| -4000- -3000 | 7 |
| -5000- -4000 | 3 |
|  | N=100 |

18. Find out the standard Deviation and its     coefficient of variation.

| x | f |
|---|---|
| 500-600 | 20 |
| 400-500 | 30 |
| 300-400 | 60 |
| 200-300 | 20 |
| 100-200 | 10 |
| 0-100 | 8 |
| -100 - 0 | 12 |
| -200- -100 | 16 |
| -300- -200 | 20 |
| | 196 |

19. Compute standard Deviation and its coefficient of variation.

| x | f |
|---|---|
| -40 – -30 | 10 |
| -30 – -20 | 28 |
| -20 – -10 | 30 |
| -10 – 0 | 42 |
| 0-10 | 65 |
| 10 –20 | 180 |
| 20 –30 | 10 |
| | 315 |

**T. Nageswara Rao**
**Vice - Principal**
**Hindu College, Guntur.**

# CO EFFICIENT OF CORRELATION

**OBJECTIVES :**

By the study of this lesson, you will be able to understand the meaning, definition of Karl Pearson's Coefficient of Correlation, probable error, uses of correlation and method of computation of coefficient of correlation.

**STRUCTURE OF THE LESSON:**

## 12.1    INTRODUCTION :

" Correlation " means a possible connection or relationship or interdependence between the values of two or more variables of the same phenomenon or individual series. It indicates the strength of the relationship. If we measure the heights and weights of 'n' individuals we assume two values - one relating to heights and the other relating to weights. Such distributions, in which each unit of the series assumes two values are called "Bivariate Distributions". If there are more than two variables in each unit such distributions are called " Multivariate Distributions " .

We can establish the relationship between the two or more values of the same series for the purpose of comparative study. Such a relationship can be established logically with some beliefs or assumptions or notions. It is purely a guess work. It does not relate to the establishment or cause and effect. However, there may or may not be the factor or causation. There may be third group of influencing factors of the changes in the values of variables. Thus sometimes, the existence of relationship is just purely a chance or accidental event.

## 12.2    TYPES OF CORRELATION :

Correlation is classified, into the following ways.

a) **Positive Correlation :** If the values of the two variables deviate in the same direction, it is said to be positive or Direct correlation.

b) **Negative Correlation :** When the values of two variables deviate in the oppositedirection, it is said to be " Negative " or " Indirect " correlation.

c) **Partial Correlation :** When one variable is independent and the other variable isdependent on the former it is said to be " Partial correlation ".

d) **Simple Correlation :** When only two variables are studied, it is called " Simple Correlation ". It means the study involves only two variables which are changing either in the same or opposite direction.

e) **Multiple Correlation :** When three or more variables are studied, it is called a "Multiple Correlation ". The variables may change in the same direction or in differentdirection.

f) **Linear Correlation :** If for corresponding to a unit change in one variable there isa constant change in the other variable over the entire range of the values it is saidto be a " Linear correlation ".

g) **Non - linear Correlation :** If the variables under study are graphed and the plottedpoints donot form a straight line. It is said to be a " Non- Linear correlation " or "Curvi- Linear correlation " The amount of change in one variable does not bear aconstant change in the other variable.

## 12.3    METHOD OF COMPUTING CORRELATION :

### Karl Pearson's Coefficient of correlation :

Karl Pearson ( 1807 - 1936 ) a great British Biometrician and statistician has propoundedthe formula for calculating the coefficient of correlation. The formula is based on arithmetic meanand Standard Deviation and it is most widely used.

The formula indicates whether the correlation is positive or negative. The answer lies between +1 and -1 (Perfect positive and Negative correlation respectively ). Zero represents the absenceof correlation. The formula is subject to algebraic manipulations and it is based on covariance is ahighly useful concept in the statistical analysis. Karl pearson's coefficient of correlation is also known as the " Product Moment Coefficient ". It is denoted by ' $\gamma$ '. It is a measure of association.

Karl Pearson's coefficient of correlation =

$$\gamma = \frac{\in dx \ dy \ x \ N - ( \in dx \ . \ \in dy)}{\sqrt{\in dx^2 x \ N - ( \in dx)^2} \ x \ \sqrt{\in dy^2 \ x \ N - ( \in dy)^2}}$$

Where

$\gamma$ = Coefficient of Correlation

$\in dx$ = Total of the deviations taken from the assumed mean in ' x ' series.

$\in dy$= Total of the deviations taken from the assumed mean in ' y' series.

$\in dx^2$ = Total of the squares of the deviations taken from the assumed mean in'x'series

$\in dy^2$ = Total of the squares of the deviations taken from the assumed mean in 'Y'series

$\in dxdy$ = Total of deviations in x & y series multiplied by each other N = Number of pairs

## 12.4  PROBABLE ERROR :

It is a difference resulting due to taking samples from the mass or population. According to "Secrist " the probable error of the correlation coefficient is an amount, which if added to and subtracted from the average correlation coefficient, produces amounts with in which the chancesare even that a coefficient of correlations from a series selected at random will fall.

With the help of probable error, it is possible to determine the reliability of the value of the coefficient in so far as it depends on the conditions of random sampling. It is an old measure of testing the reliability of an observed value of correlation coefficient. It is based on the standard errors multiplied by the probable error. It is obtained by the formula.

## 12.5   EXAMPLES :

### Example 1 :

From the following data compute Karl Pearson's coefficient of correlationWages  :

| | 100 | 101 | 102 | 102 | 100 | 99 | 97 | 98 | 96 | 95 |
|---|---|---|---|---|---|---|---|---|---|---|
| Cost of living  : | 98 | 99 | 99 | 97 | 95 | 92 | 95 | 94 | 90 | 91 |

### Solution :

| X | dx | dx² | y | dy | dy² | dxdy |
|---|---|---|---|---|---|---|
| 100 | 0 | 0 | 98 | +3 | 9 | 0 |
| 101 | +1 | 1 | 99 | +4 | 16 | +4 |
| 102 | +2 | 4 | 99 | +4 | 16 | +8 |
| 102 | +2 | 4 | 97 | +2 | 4 | +4 |
| 100 | 0 | 0 | 95 | 0 | 0 | 0 |
| 99 | -1 | 1 | 92 | -3 | 9 | +3 |
| 97 | -3 | 9 | 95 | 0 | 0 | 0 |
| 98 | -2 | 4 | 94 | -1 | 1 | +2 |
| 96 | -4 | 16 | 90 | -5 | 25 | +20 |
| 95 | -5 | 25 | 91 | -4 | 16 | +20 |
| N=10 | -10 | 64 | N=10 | 0 | 96 | 61 |

$$\gamma = \frac{\in dx\, dy \times N - (\in dx \cdot \in dy)}{\sqrt{\in dx^2 \times N - (\in dx)^2} \times \sqrt{\in dy^2 \times N - (\in dy)^2}}$$

$$= \frac{61\times10 - (-10\times0)}{\sqrt{64\times10 - (-10)^2} \times \sqrt{96\times10 - (0)^2}}$$

$$= \frac{610 - 0}{\sqrt{640 - 100} \times \sqrt{960}}$$

$$= \frac{610}{\sqrt{540} \times \sqrt{960}}$$

$$= \frac{610}{23.2379 \times 30.9838}$$

$$= \frac{610}{719.99}$$

$$= +0.8472 \text{ Positive.}$$

**Example 2**

Compute Karl Pearson's coefficient of correlation from the following data.

| x : | 27 | 21 | 35 | 44 | 29 | 30 | 32 | 42 | 41 | 36 | 28 | 26 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| y : | 40 | 37 | 21 | 25 | 36 | 41 | 22 | 31 | 23 | 24 | 39 | 37 |

**Solution :**

| X | dx dx² | | y | dy | dy²dxdy |
|---|---|---|---|---|---|
| 27 | -3 | 9 | 40 | +981 | -27 |
| 21 | -981 | | 37 | +636 | -54 |
| 35 | +525 | | 21 | -10100 | -50 |
| 44 | +14196 | | 25 | -636 | -84 |
| 29 | -1 | 1 | 36 | -525 | -5 |
| 30 | 0 | 0 | 41 | +10100 | 0 |
| 32 | +2 | 4 | 22 | -981 | -18 |
| 42 | +12144 | | 31 | 00 | 0 |
| 41 | +11121 | | 23 | -864 | -88 |
| 36 | +636 | | 24 | -749 | -42 |
| 28 | -2 | 4 | 39 | +864 | -16 |
| 26 | 416 | | 37+6 | 36 | -24 |
| N=12 | 31 637 | | N=12 | 4 | 672 -408 |

$$\gamma = \frac{\in dx\, dy \times N - (\in dx \,.\, \in dy)}{\in dx^2 \times N - (\in dx)^2 \quad \times \in dy^2 \times N - (\in dy)^2}$$

$$= \frac{-408 \times 12 - (31 \times 4)}{\sqrt{637 \times 12 - (31)^2} \times \sqrt{672 \times 12 - (4)^2}}$$

$$= \frac{-4896 - 124}{\sqrt{7644 - 961} \times \sqrt{8064 - 16}}$$

$$= \frac{-5020}{81.7496 \times 89.7106}$$

$$= \frac{-5020}{7333.81} = -0.6845 \text{ Negative.}$$

**Example 3 :**

Compute Karl Pearson's coefficient of correlation from the following data.

x :  300    350    400    450    500    550    600    650    700

y :  800         900 1000 1100 1200 1300 1400 1500 1600

**Solution :**

| X | dx | dx² | y | dy | dy² | dxdy |
|---|---|---|---|---|---|---|
| 300 | -4 | 16 | 800 | -4 | 16 | 16 |
| 350 | -3 | 9 | 900 | -3 | 9 | 9 |
| 400 | -2 | 4 | 1000 | -2 | 4 | 4 |
| 450 | -1 | 1 | 1100 | 1 | 1 | 1 |
| 500 | 0 | 0 | 1200 | 0 | 0 | 0 |
| 550 | +1 | 1 | 1300 | +1 | 1 | 1 |
| 600 | +2 | 4 | 1400 | +2 | 4 | 4 |
| 650 | +3 | 9 | 1500 | +3 | 9 | 9 |
| 700 | +4 | 16 | 1600 | +4 | 16 | 16 |
| N=9 | 0 | 60 | N=9 | 0 | 60 | 60 |

$$\gamma = \frac{\in dx\ dy \times N - (\in dx \cdot \in dy)}{\sqrt{\in dx^2 \times N - (\in dx)^2} \times \sqrt{\in dy^2 \times N - (\in dy)^2}}$$

$$= \frac{60 \times 9 - (0 \times 0)}{\sqrt{60 \times 9 - (0)^2} \times \sqrt{60 \times 9 - (0)^2}}$$

$$= \frac{540}{\sqrt{540} \times \sqrt{540}}$$

$= \dfrac{540}{540}$

$= + 1$ Positive

**Example 4 :**

Compute Karl Pearson's coefficient of correlation

x :                 20    40    60    80   100   120   140   160   180  200

y : 2000 1990 1980 1970 1960 1950 1940 1930 1920 1910

**Solution :**

| X | dx | dx² | | y | dy | dy² | dxdy |
|---|----|----|---|---|----|----|------|
| 20 | -4 | 16 | | 2000 | | +525 | -20 |
| 40 | -3 | 9 | | 1990 | | +416 | -12 |
| 60 | -2 | 4 | | 1980 | | +39 | -6 |
| 80 | -1 | 1 | | 1970 | | +24 | -2 |
| 100 | 0 | 0 | | 1960 | | +11 | 0 |
| 120 | +1 | 1 | | 1950 | | 00 | 0 |
| 140 | +2 | 4 | | 1940 | | -11 | -2 |
| 160 | +3 | 9 | | 1930 | | -24 | -6 |
| 180 | +4 | 16 | | 1920 | | -39 | -12 |
| 200 | +5 | 25 | 1910 | | -4 | 16 | -20 |
| N=10 | 5 | 85 | | N=10 | | 5   85 | -80 |

$$\dfrac{\in dx\ dy \ x \ N - (\in dx \ . \ \in dy)}{\sqrt{\in dx^2 x \ N - (\in dx)^2} \ x \ \sqrt{\in dy^2 \ x \ N - (\in dy)^2}}$$

$= \dfrac{-80 \times 10 - (5 \times 5)}{\sqrt{85 \times 10 - (5)^2} \ x \ \sqrt{85 \times 10 - (5)^2}}$

$= \dfrac{-800 - 25}{\sqrt{850 - 25} \ x \sqrt{850 - 25}}$

$$= \frac{-825}{\sqrt{825} \times \sqrt{825}} \qquad = \frac{-825}{825}$$

= -1 Negative

**Example 5 :**

Compute Karl Pearson's coefficient of correlation and probable Error.

| Marks in Accounts: | 50 | 60 | 58 | 47 | 49 | 33 | 65 | 43 | 46 | 68 |
|---|---|---|---|---|---|---|---|---|---|---|
| Marks in Q.T. : | 48 | 65 | 50 | 48 | 55 | 58 | 63 | 48 | 50 | 70 |

**Solution :**

| X | dx | dx² | y | dy | dy² | dxdy |
|---|---|---|---|---|---|---|
| 50 | 0 | 0 | 48 | | -749 | 0 |
| 60 | +10 | 100 | 65 | | +10100 | 100 |
| 58 | +8 | 64 | 50 | | -525 | -40 |
| 47 | -3 | 9 | 48 | | -749 | 21 |
| 49 | -1 | 1 | 55 | | 00 | 0 |
| 33 | -17 | 289 | 58 | | +39 | -51 |
| 65 | +15 | 225 | 63 | | +864 | -120 |
| 43 | -7 | 49 | 48 | | -749 | 49 |
| 46 | -4 | 16 | 50 | | -525 | 20 |
| 68 | 18 | 324 | 70 | +20 | 400 | 360 |
| N=10 | 19 | 1077 | N=10 | | 10770 | 579 |

$$\gamma = \frac{\in dx\ dy \times N - (\in dx\ .\ \in dy)}{\sqrt{\in dx^2 \times N - (\in dx)^2} \times \sqrt{\in dy^2 \times N - (\in dy)^2}}$$

$$= \frac{579 \times 10 - (19 \times 10)}{\sqrt{1077 \times 10 - (19)^2} \times \sqrt{770 \times 10 - (10)^2}}$$

$$= \frac{5790 - 190}{\sqrt{10770 - 361} \times \sqrt{7700 - 100}}$$

$$= \frac{5600}{\sqrt{10409} \times \sqrt{7600}}$$

$$\frac{5600}{102.02 \times 87.18}$$

$$\frac{5600}{8894.28}$$

$= 0.6745 \ (0.19)$

$= \pm \ 0.128$

**Example 6 :**

The population and the number of persons partially or fully blind are given in the following table . Find out whether there is any correlation between their age and their blindness.

| AgePopulation in ' 000 | | No. of persons blind |
|---|---|---|
| 0-10 | 100 | 55 |
| 10-20 | 60 | 40 |
| 20-30 | 40 | 40 |
| 30-40 | 36 | 40 |
| 40-50 | 24 | 36 |
| 50-60 | 11 | 22 |
| 60-70 | 6 | 18 |
| 70-80 | 3 | 15 |

**Solution :**

| X | Mv | dx | dx² | y | dy | dy² | dxdy |
|---|---|---|---|---|---|---|---|
| 0-10 | 5 | -416 | | 55-45 | 2025 | | 180 |
| 10-20 | 15 | -39 | | 67-33 | 1089 | | 99 |
| 20-30 | 25 | -24 | | 1000 | | 0 | 0 |
| 30-40 | 35 | -11 | | 11111 | 121 | | -11 |
| 40-50 | 45 | 00 | | 15050 | 2500 | | 0 |
| 50-60 | 55 | +11 | | 200100 | 10000 | | 100 |
| 60-70 | 65 | +24 | | 300200 | 40000 | | 400 |
| 70-80 | 75 | +3 | 9500 | 400 | 160000 | | 1200 |
| | N=8 | -4 44 | | N=8683 | 215735 | | 1968 |

$$\gamma = \frac{\in dx\ dy \times N - (\in dx\ .\ \in dy)}{\sqrt{\in dx^2 \times N - (\in dx)^2} \times \sqrt{\in dy^2 \times N - (\in dy)^2}}$$

$$= \frac{1968 \times 8 - (-4 \times 683)}{\sqrt{44 \times 8 - (-4)^2} \times \sqrt{215735 \times 8 - (683)^2}}$$

$$= \frac{15744 + 2732}{\sqrt{352 - 16} \times \sqrt{1725880 - 466489}}$$

$$= \frac{18476}{\sqrt{336} \times \sqrt{1259391}}$$

$$= \frac{18476}{18.33 \times 1122.23}$$

$$= \frac{18476}{20570.48}$$

= +0.898 Positive

Thus it can be said that there is correlation between the age and blind ness.

## 12.6  MERITS OF CO EFFICIENT OF CORRELATION :

1. **Counts all values :** It takes into account all values of the given data of x & y. Therefore it is based on all observations of the series.

2. **More practical and popular :** Karl Pearson's correlation is considered to be more practical

method as compared to other mathematical methods used for ' $\gamma$ '. It is also very popular and as such commonly used method.

3. **Numerical measurement of ' $\gamma$ ' :** It provides numerical measurement of Coefficient of correlation.

4. **Measures degree and direction :** This method measures both degree and direction of the correlation between the variables at a time.

5. **Facilitates comparison :** It is a pure number independent of units. Therefore thecomparison between the series can be done easily.

6. **Algebraic treatment possible :** This technique can be easily applied for higher algebraic treatment.

## 12.7 DEMERITS OF CO EFFICIENT OF CORRELATION :

1. **Linear relationship :** It assumes linear relationship between the variables regardless of the fact whether that assumption is correct or not.

2. **More time consuming :** Compared with some other methods, this method ismore time consuming.

3. **Affected by extreme items :** This method is affected by extreme items.

4. **Difficult to interpret :** It is not easy to interpret the significance of correlationefficient. It is generally misinterpreted.

## 12.8 SUMMARY :

Karl Pearson's coefficient of correlation method gives a precise and summary quantitativefigure which can be meaningfully interpreted. It gives either positive or negative direction or degreeof the relationship between the two variables.

## 12.9 QUESTIONS :

1. What is meant by coefficient of correlation ?

2. State the types of correlation

3. Explain the method of computing Coefficient of correlation

4. Explain about the probable error.

5. What are the merits and demerits of Co efficient of correlation ?

6. State the assumption of Karl Pearson's Co efficient of correlation ?

7. What is meant by Linear and non- linear correlation ?

## 12.10 EXERCISES :

1. Compute Karl Pearson's Co efficient of correlation

Age of Husband :25    22    28    26    35    20    22    40    20    18    19    25

Age of Wife : 18    15    20    17    22    14    16    21    15    14    15    17

2.     Find out  Karl Pearson's Co efficient of correlation and Probable Error.

Age of Husband : 23     27     28     29     30     31     33     35     36     39

Age of Wife :    18     22     23     24     25     26     28     29     30     32

3.  Ascertain Karl Pearson's Co efficient of correlation and Probable Error.

x :     25  22   28   26   35   20   22   40   20   18   19   25

y :     18  15   20   17   22   14   16   21   15   14   15   17

4.  Compute Karl Pearson's Co efficient of correlation and Probable Error.

x :     10  12   18   16   15   19   18   17   15   16

y:      30  35   45   44   42   48   47   46   44   45

5.  In the following data, the population and the number persons partly or fully deaf, aregiven. Find out whether there is any relationship between their age and deafness.

| Age | Population in thousands | No.of persons deal |
|-----|-------------------------|--------------------|
| 0-10  | 100 | 60 |
| 10-20 | 80  | 45 |
| 20-30 | 60  | 43 |
| 30-40 | 40  | 42 |
| 40-50 | 30  | 33 |
| 50-60 | 20  | 30 |
| 60-70 | 10  | 26 |
| 70-80 | 5   | 24 |

6.     Find out the Co efficient of correlation between the following two variables. Comment onthe result through the Probable Error.

(x) :    6         8   12   15        18   20        24   28   31

(f) :    10        12  15   15        18   25        22   26   28

7.     Calculate the Co efficient of correlation from the following data and calculate ProbableError.

Q.T. (x) :     30     60     30     66     72     24     18     12     42     6

Accounts (f) : 06     36     12     48     30     06     24     36     30     12

8. Calculate the Co efficient of correlation between income and weight from the followingdata. Comment on the result.

| Income (Rs.) : | 100 | 200 | 300 | 400 | 500 | 600 |
|---|---|---|---|---|---|---|
| Weight (lbs) : | 120 | 130 | 140 | 150 | 160 | 170 |

9. Calculate Karl Pearson's Co efficient of correlation from the following data

| (x) : | 150 | 200 | 250 | 300 | 350 | 400 | 450 | 500 | 550 | 600 |
|---|---|---|---|---|---|---|---|---|---|---|
| (f) : | 600 | 575 | 550 | 525 | 500 | 475 | 450 | 425 | 400 | 375 |

10. Calculate Karl Pearson's Co efficient of correlation of x and y variables.

| (x) : | 15 | 18 | 30 | 27 | 25 | 23 | 30 |
|---|---|---|---|---|---|---|---|
| (f) : | | 7 | 10 | 17 | 16 | 12 | 13 | 9 |

11. Compute Karl Pearson's Co efficient of correlation for the following data( Height in inches)

| of Husband x : | 60 | 62 | 64 | 66 | 68 | 70 | 72 |
|---|---|---|---|---|---|---|---|
| of Wife  Y : | 61 | 63 | 63 | 63 | 64 | 65 | 67 |

12. Calculate Karl Pearson's Co efficient of correlation from the following data

| (x) : | 12 | 9 | 8 | 10 | 11 | 13 | 7 |
|---|---|---|---|---|---|---|---|
| (f) : | | 14 | 8 | 6 | 9 | 11 | 12 | 3 |

13. Find out Co efficient of correlation from the following data

| (x) : | 3 | 5 | 6 | 7 | 9 | 12 |
|---|---|---|---|---|---|---|
| (f) : | | 20 | 14 | 12 | 10 | 9 | 7 |

14. Calculate the Co efficient of correlation between Advertisement cost and sales as perthe data given below :

| Cost in thousands : | 39 | 65 | 62 | 90 | 82 | 75 | 25 | 98 | 36 | 78 |
|---|---|---|---|---|---|---|---|---|---|---|
| Sales in Lakhs: | 47 | 53 | 58 | 86 | 62 | 68 | 60 | 91 | 51 | 84 |

**T. Nageswara Rao**
**Vice - Principal**
**Hindu College, Guntur.**

## LESSON – 13

# MEASURES OF SKEWNESS AND KURTOSIS

**LEARNING OBJECTIVES :**

✓     To distinguish between a symmetrical and a skewed distribution;

✓     To compute various coefficients to measure the extent of skewness in a distribution;

✓     To distinguish between platykurtic, mesokurtic and leptokurtic distributions; and compute the coefficient of kurtosis.

**STRUCTURE OF THE LESSON :**

## 13.1 INTRODUCTION :

Skewness means lack of symmetry. In mathematics, a figure is called symmetric if there exists a point in it through which if a perpendicular is drawn on the X-axis, it divides the figure into two congruent parts i.e. identical in all respect or one part can be superimposed on the other i.e mirror images of each other. In Statistics, a distribution is called symmetric if mean, medianand mode coincide. Otherwise, the distribution becomes asymmetric. If the right tail is longer,we get a positively skewed distribution for which mean > median > mode while if the left tail is longer, we get a negatively skewed distribution for which mean < median < mode.

### 13.2 MEANING OF SKEWNESS :

- Skewness means lack of symmetry.
- We study skewness to have an idea about the shape of the curve which we can draw with the help of the given data.
- If, in a distribution, Mean = Median = Mode, then that distribution is known as Symmetrical Distribution.

If, in a distribution, Mean ≠ Median ≠ Mode, then it is not a symmetrical distribution and it is called a Skewed Distribution and such a distribution could either be positively skewed or negatively skewed.

The example of the Symmetrical curve, Positive skewed curve and Negative skewed curve are given as follows:
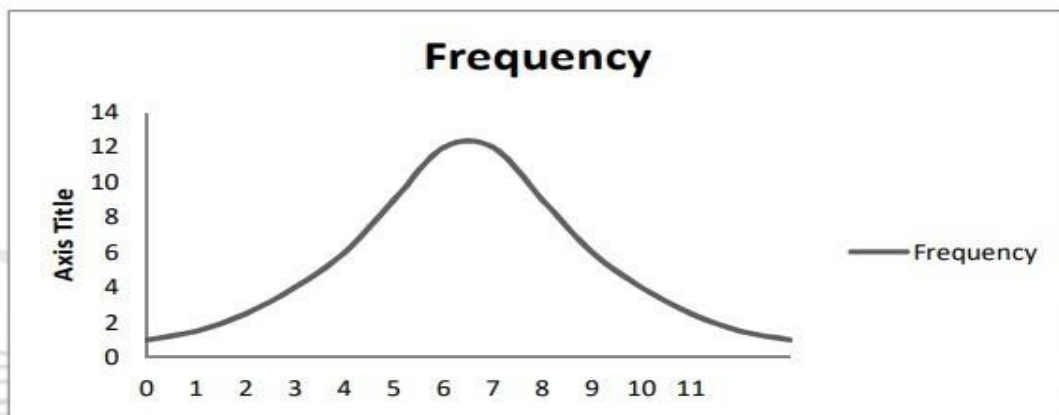


**Symmetrical Curve**

It is clear from the above diagram, in a negatively skewed distribution, the value of the mode is maximum and that of the mean is least. The median lies in between the two. In the negatively skewed distribution the frequencies are spread out over a greater range of values on the left hand side than they are on the right hand side
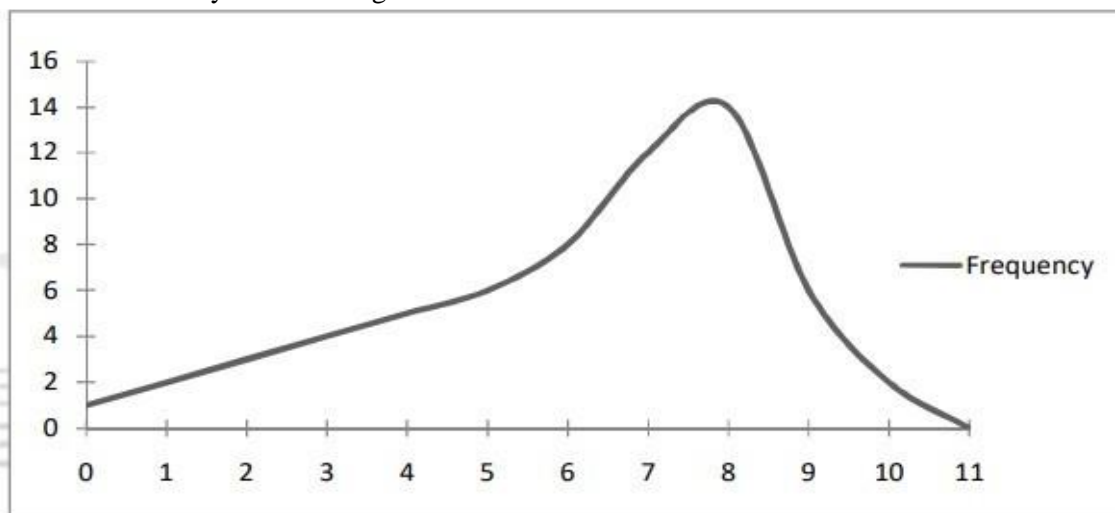


**Negative Skewed Curve**

It is clear from the above diagram that in a symmetrical distribution the values of mean, median and mode coincide. The spread of the frequencies is the same on both sides of the center point of the curve.
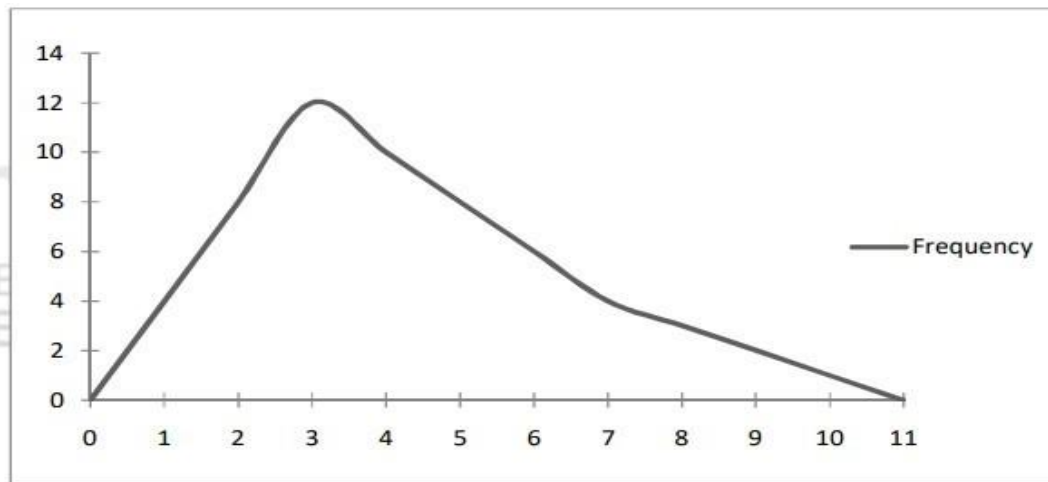


**Positive Skewed Curve**

## 13.3  DIFFERENCE BETWEEN VARIANCE AND SKEWNESS :

The following two points of difference between variance and skewness should be carefully noted.

1.    Variance tells us about the amount of variability while skewness gives the direction of variability.

2.    In business and economic series, measures of variation have greater practical application than measures of skewness. However, in medical and life science field measures of skewness have greater practical applications than the variance.

## 13.4  VARIOUS MEASURES OF SKEWNESS :

Measures of skewness help us to know to what degree and in which direction (positive or negative) the frequency distribution has a departure from symmetry. Although positive or negative skewness can be detected graphically depending on whether the right tail or the left tail is longer but, we don't get idea of the magnitude. Besides, borderline cases between symmetry and asymmetry may be difficult to detect graphically. Hence some statistical measures are required to find the magnitude of lack of symmetry. A good measure of skewness should possessthree criteria :

1.    It should be a unit free number so that the shapes of different distributions, so far as symmetry is concerned, can be compared even if the unit of the underlying variables are different;

2.    If the distribution is symmetric, the value of the measure should be zero. Similarly, the measure should give positive or negative values according as the distribution has positive or negative skewness respectively; and

3.  As we move from extreme negative skewness to extreme positive skewness, the value of the measure should vary accordingly.

Measures of skewness can be both absolute as well as relative. Since in a symmetrical distribution mean, median and mode are identical more the mean moves away from the mode, the larger the asymmetry or skewness. An absolute measure of skewness can not be used for purposes of comparison because of the same amount of skewness has different meanings in distribution with small variation and in distribution with large variation.

### 13.4.1  Absolute Measures of Skewness :

Following are the absolute measures of skewness :

1.  Skewness (Sk) = Mean – Median

2.  Skewness (Sk) = Mean – Mode

3.  Skewness (Sk) = (Q3 - Q2) - (Q2 - Q1)

For comparing to series, we do not calculate these absolute measures we calculate the relative measures which are called coefficient of skewness.

Coefficient of skewness are pure numbers independent of units of measurements.

### 13.4.2  Relative Measures of Skewness :

In order to make valid comparison between the skewness of two or more distributions we have to eliminate the distributing influence of variation. Such elimination can be done by dividing the absolute skewness by standard deviation. The following are the important methods of measuring relative skewness:

### 13.4.3  Karl Pearson  Coefficient of Skewness :

Karl Pearson defined the following coefficients of skewness, based upon the second and third central moments
According to Karl – Pearson, the absolute measure of skewness = mean – mode. This measure is not suitable for making valid comparison of the skewness in two or more distributions because the unit of measurement may be different in different series

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

It is used as measure of skewness. For a symmetrical distribution, $\beta_1$ shall be zero. $\beta_1$ as a measure of skewness does not tell about the direction of skewness, i.e. positive or negative. Because $\mu_3$ being the sum of cubes of the deviations from mean may be positive or negative but $\mu_3{}^2$ is always positive. Also, $\mu_2$ being the variance always positive. Hence, $\beta_1$ would be always positive. This drawback is removed if we calculate Karl Pearson's Gamma coefficient $\gamma_1$ which is the square root of $\beta_1$ i. e.

$$\gamma_1 = \pm \sqrt{\beta_1} = \frac{\mu_3}{\left(\mu_2\right)^{3/2}} = \frac{\mu_3}{\sigma^3}$$

Then the sign of skewness would depend upon the value of $\mu_3$ whether it is positive or negative. It is advisable to use $\gamma_1$ as measure of skewness.
This method is most frequently used for measuring skewness. The formula for measuring coefficient of skewness is given by

$$S_k = \frac{Mean - Mode}{\sigma}$$

The value of this coefficient would be zero in a symmetrical distribution. If mean is greater than mode, coefficient of skewness would be positive otherwise negative. The value of the Karl Pearson's coefficient of skewness usually lies between □1 for moderately skeweddistribution. If mode is not well defined, we use the formula

$$S_k = \frac{3(Mean - Median)}{\sigma}$$

By using the relationship

Mode = (3 Median – 2 Mean)

Here, $-3 \le S_k \le 3$. In practice it is rarely obtained.

### 13.4.4 Bowleys's Coefficient of Skewness :

This method is based on quartiles. The formula for calculating coefficient of skewness is given by Bowleys's Coefficient of Skewness This method is based on quartiles. The formula for calculating coefficient of skewness is given by

$$S_k = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)}$$
$$= \frac{(Q_3 - 2Q_2 + Q_1)}{(Q_3 - Q_1)}$$

The value of $S_k$ would be zero if it is a symmetrical distribution. If the value is greater than zero, it is positively skewed and if the value is less than zero it is negatively skewed distribution. It will take value between +1 and -1

### 13.4.5 Kelly's Coefficient of Skewness :

The coefficient of skewness proposed by Kelly is based on percentiles and deciles. The formula for calculating the coefficient of skewness is given by

**Based on Percentiles**

$$S_k = \frac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{(P_{90} - P_{10})}$$
$$= \frac{(P_{90} - 2P_{50} + P_{10})}{(P_{90} - P_{10})}$$

where, $P_{90}$, $P_{50}$ and $P_{10}$ are 90th, 50th and 10th Percentiles.

**Based on Deciles**

$$S_k = \frac{(D_9 - 2D_5 + D_1)}{D_9 - D_1}$$

where, $D_9$, $D_5$ and $D_1$ are 9th, 5th and 1st Decile.

Example1 : For a distribution Karl Pearson's coefficient of skewness is 0.64, standard deviationis 13 and mean is 59.2 Find mode and median.

Solution : We have given

$$S_k = 0.64, \sigma = 13 \text{ and Mean} = 59.2$$

Therefore by using formulae

$$S_k = \frac{\text{Mean} - \text{Mode}}{\sigma}$$

$$0.64 = \frac{59.2 - \text{Mode}}{13}$$

$$\text{Mode} = 59.20 - 8.32 = 50.88$$

$$\text{Mode} = 3 \text{ Median} - 2 \text{ Mean}$$

$$50.88 = 3 \text{ Median} - 2 (59.2)$$

$$\text{Median} = \frac{50.88 + 118.4}{3} = \frac{169.28}{3} = 56.42$$

## 13.5 CONCEPT OF KURTOSIS :

If we have the knowledge of the measures of central tendency, dispersion and skewness, even then we cannot get a complete idea of a distribution. In addition to these measures, we need to know another measure to get the complete idea about the shape of the distribution which can be studied with the help of Kurtosis. Prof. Karl Pearson has called it the "Convexity of a Curve". Kurtosis gives a measure of flatness of distribution.

The degree of kurtosis of a distribution is measured relative to that of a normal curve. The curves with greater peaked ness than the normal curve are called "Leptokurtic". The curves which are more flat than the normal curve are called "Platykurtic". The normal curve is called "Mesokurtic." The Fig.4 describes the three different curves mentioned above :
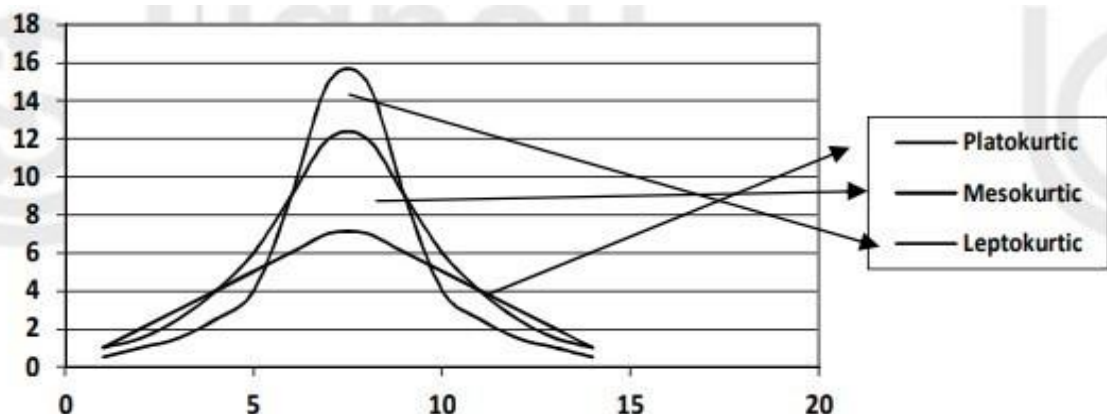


**Platykurtic Curve, Mesokurtic Curve and Leptokurtic Curve**

## 13.6 MEASURES OF KURTOSIS :

### 13.6.1 Karl Pearson's Measures of Kurtosis :

For calculating the kurtosis, the second and fourth central moments of variable are used. For this, following formula given by Karl Pearson is used :

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

or $\qquad \gamma_2 = \beta_2 - 3$

where, $\mu_2$ = Second order central moment of distribution

$\mu_4$ = Fourth order central moment of distribution

where, $\mu_2$ = Second order central moment of distribution

$\mu_4$ = Fourth order central moment of distribution

### Description:

1. If $\beta_2 = 3$ or $\gamma_2 = 0$, then curve is said to be mesokurtic;

2. If $\beta_2 < 3$ or $\gamma_2 < 0$, then curve is said to be platykurtic;

3. If $\beta_2 > 3$ or $\gamma_2 > 0$, then curve is said to be leptokurtic;

### 13.6.2 Kelly's Measure of Kurtosis :

Kelly has given a measure of kurtosis based on percentiles. The formula is given by

$$\beta_2 = \frac{P_{75} - P_{25}}{P_{90} - P_{10}}$$

where, $P_{75}$, $P_{25}$, $P_{90}$, and $P_{10}$ are $75^{th}$, $25^{th}$, $90^{th}$ and $10^{th}$ percentiles of dispersion respectively.

If $\beta_2 > 0.26315$, then the distribution is platykurtic.

If $\beta_2 < 0.26315$, then the distribution is leptokurtic.

Example 2 : First four moments about mean of a distribution are 0, 2.5, 0.7 and 18.75. Find coefficient of skewness and kurtosis.

**Solution:** We have $\mu_1 = 0$, $\mu_2 = 2.5$, $\mu_3 = 0.7$ and $\mu_4 = 18.75$

Therefore, Skewness, $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = \dfrac{(0.7)^2}{(2.5)^3} = 0.031$

Kurtosis, $\beta_2 = \dfrac{\mu_4}{\mu_2^2} = \dfrac{18.75}{(2.5)^2} = \dfrac{18.75}{6.25} = 3.$

As $\beta_2$ is equal to 3, so the curve is mesokurtic.

**E1)** Using the formulae, we have

$$S_k = \frac{Mean - Mode}{\sigma}$$

$$1.28 = \frac{164 - 100}{\sigma}$$

$$\sigma = \frac{64}{1.28} = 50$$

**E2)** We have given $S_k = 1.2$

$$Q_1 + Q_3 = 200$$
$$Q_2 = 76$$

then $S_k = \dfrac{(Q_3 + Q_1 - 2Q_2)}{(Q_3 - Q_1)}$

$$1.2 = \frac{(200 - 2 \times 76)}{(Q_3 - Q_1)}$$

$$Q_3 - Q_1 = \frac{48}{1.2} = 40$$

$$Q_3 - Q_1 = 40$$

4. Calculate Karl Pearson's coefficient of Skewness for the following data.

| Wage per Item Rs.(x) | Number of items f | fx | $x^2$ | $fx^2$ |
|---|---|---|---|---|
| 12 | 10 | 120 | 144 | 1440 |
| 15 | 25 | 375 | 225 | 5625 |
| 20 | 40 | 800 | 400 | 16000 |
| 25 | 70 | 1750 | 625 | 43750 |
| 30 | 32 | 960 | 900 | 28800 |
| 40 | 13 | 520 | 1600 | 20800 |
| 50 | 10 | 500 | 2500 | 25000 |
| | $\sum f = 200$ | $\sum fx = 5025$ | | $\sum fX^2 = 141415$ |

$$\overline{X} = \frac{\sum fX}{\sum f} = \frac{5025}{200} = 25.13$$

$$\sigma = \sqrt{\frac{\sum fX^2}{\sum f} - \left[\frac{\sum fX}{\sum f}\right]^2} =$$

$$\sqrt{\frac{141415}{200} - (25.13)^2} =$$

$$\sqrt{707.075 - 631.5169} = \sqrt{75.5581}$$

$$= 8.69$$

Greatest frequency = 70,      $Z = 25$

$$Sk_p = \frac{\overline{X} - Z}{\sigma} = \frac{25.13 - 25}{8.69} = 0.13/8.69 = 0.0149$$

5. Calculate Karl Pearson's coefficient of Skewness for the following data.

| Profit (Rs.Lakhs) | No of Companies (f) | m | fm | $m^2$ | $fm^2$ |
|---|---|---|---|---|---|
| 10-20 | 18 | 15 | 270 | 225 | 4050 |
| 20-30 | $20 = f_0$ | 25 | 500 | 625 | 12500 |
| 30-40 | $30 = f_1$ | 35 | 1050 | 1225 | 36750 |
| 40-50 | $22 = f_2$ | 45 | 990 | 2025 | 44550 |
| 50-60 | 10 | 55 | 550 | 3025 | 30250 |
| | $\sum f = 100$ | | $\sum fm = 3360$ | | $\sum fm^2 = 128100$ |

$$\overline{X} = \frac{\sum fm}{\sum f} = 3360/100 = 33.6$$

$$\sigma = \sqrt{\frac{\sum fm^2}{\sum f} - \left[\frac{\sum fm}{\sum f}\right]^2} = \sqrt{\frac{128100}{100} - (33.6)^2} = \sqrt{1281 - 1128.96} = \sqrt{152.04} = 12.33$$

$D_1 = f_1 - f_0 = 30 - 20 = 10$: $D_2 = f_1 - f_2 = 30 - 22 = 8$: L = 30: i=10

$$Z = L + \left[\frac{D_1}{D_1 + D_2}\right]i = 30 + \left[\frac{10}{10+8}\right]10 = 30 + \left[\frac{10}{18}\right]10 = 30 + 5.56 = 35.56$$

$$Sk_p = \frac{\overline{X} - Z}{\sigma} = \frac{33.6 - 35.56}{12.33} = -1.96/12.33 = -0.1590$$

## 13.7 SUMMARY :

A fundamental task in many statistical analyses is to characterize the location and variability of a data set. A further characterization of the data includes skewness and kurtosis. Measure of Dispersion tells us about the variation of the data set. Skewness tells us about the direction of variation of the data set.

## 13.8 KEY WORDS :

**Skewness :** Departure from symmetry is skewness.

**Moment of Order r** : It is defined as the arithmetic mean of the rth powerofdeviations of observations.

**Coefficient of Kurtosis** : It is a measure of the relative peaked ness of the top of a frequency curve.

## 13.9 SELF ASSESSMENT QUESTIONS :

1. Discuss the types of skewness

2. Examine the difference between the skewness and kurtosis

3. Describe the Measures of Kurtosis

## 13.10 SUGGESTED READINGS :

1. Ranjit Kumar Research Methodology, Sage Publications 2018.

2. Deepak Chawla, Research Methodology concepts and cases, 2 nd Edtion 2022.

3. M.V.Kulkarni , Research Methodology, Everest Publishing House, 2022

4. Dr.Vijaya upagade, Research Methodology, S.Chand Publications 2022

**Dr. V. Naga Nirmala**

# SPEARMAN'S RANK CORRELATION

**OBJECTIVES :**

By the study of this lesson, you will be able to understand, the meaning definition and uses of Spearman's Rank correlation with examples.

### STRUCTURE OF THE LESSON :

## 14.1    INTRODUCTION :

Charles Edward spearman, a British Psychologist, developed a formula to obtain the rank correlation coefficient in 1904. He has tried to establish the rank correlation coefficient between the Ranks of 'n' individuals in the two or more variables ' Accordingly, it is possible for a class teacher to arrange his students in an ascending order or in descending order of intelligence though intelligence cannot be measured quantitatively. In a similar way ranking can be made in a beauty contest and correlation can be established among the scores given by the different judge or selectors.

It is, however, possible to measure the degree of correlation between two sets of observations or between paired values when only the relative order of magnitude is given for each series. For example, suppose 10 students have appeared for two papess in a test and from actual marks obtained by them, their rankings can be determined. If we want to know whether their performances are correlated, we can use " Sperman's Rank corelation Coefficient " method. The formula is based on the ranks of the variables according to their sizes.

## 14.2    CIRCUMSTANCES WHEN THE RANK CORRELATION IS USED :

Following are the circumstances when the Rank Correlation coefficient is used.

In a beauty contest, cooking contest, flower show contest and interview involving selections, we can use the rank correlation coefficient.

If the data are irregular or extreme items are erratic or in accurate, we can usethe rank correlation coefficient.

In spearmen's coefficient of correlation we take the differences in Ranks, squaring them and finding out the aggregate of the squared differences. Symbolically.

$$\gamma_s = 1 - \frac{6 \in D^2}{N(N^2 - 1)}$$

Where

$\gamma_s$ = Coefficient of Correlation

$\in D^2$ = Total of the deviations between x & y itemsN = No. of pairs.

## 14.3 TYPES OF RANK CORRELATION :

In Rank Coefficient of correlation three different cases must be studied.

**Case I** When Ranks are not given**Case II**

When Ranks are given **Case III** When

Ranks are equal.

**When Ranks are not given :**

 **Example 1 :**

Compute Rank Correlation from the following data

x : 415 434 420 430 424 428

y : 330 332 328 331 327 325

 **Solution :**

| X | $R_1$ | y | $R_2$ | D ($R_1 - R_2$) | -$D^2$ |
|---|---|---|---|---|---|
| 415 | 6 | 330 | 3 | 3 | 9 |
| 434 | 1 | 332 | 1 | 0 | 0 |
| 420 | 5 | 328 | 4 | 1 | 1 |
| 430 | 2 | 331 | 2 | 0 | 0 |
| 424 | 4 | 327 | 5 | -1 | 1 |
| 428 | 3 | 325 | 6 | -3 | 9 |
| N=6 | | N=6 | | | $\in D^2$= 20 |

$$\gamma_s = 1 - \frac{6 \in D^2}{N(N^2-1)}$$

$$= 1 - \frac{6(20)}{6(6^2-1)}$$

$$= 1 - \frac{120}{210}$$

$$= 1 - 0.571$$

$$= 0.429$$

**When Ranks are given :**

**Example 2 :**

Compute Rank Correlation from the following data

x :   415   434   420   430   424   428

y :   330   332   328   331   327   325

**Solution :**

| X | R₁ | y | R₂ | D (R₂ – R₁) | D² |
|---|---|---|---|---|---|
| 415 | 6 | 330 | 3 | 3 | 9 |
| 434 | 1 | 332 | 1 | 0 | 0 |
| 420 | 5 | 328 | 4 | 1 | 1 |
| 430 | 2 | 331 | 2 | 0 | 0 |
| 424 | 4 | 327 | 5 | -1 | 1 |
| 428 | 3 | 325 | 6 | -3 | 9 |

$$\in D^2 = \underline{20}$$

$$\gamma_s = 1 - \frac{6 \in D^2}{N(N^2-1)}$$

$$= 1 - \frac{6(20)}{6(6^2-1)}$$

$$= 1 - \frac{120}{210} = 1 - 0.571$$

$$= 0.429$$

**Example 3 :**

The Ranks given by 3 judges to 10 participants in a beauty contest were as under.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Judge A : | 1 | 6 | 5 | 10 | 3 | 2 | 4 | 9 | 7 | 8 |
| Judge B : | 3 | 5 | 8 | 4 | 7 | 10 | 2 | 1 | 6 | 9 |
| Judge C : | 6 | 4 | 9 | 8 | 1 | 2 | 3 | 10 | 5 | 7 |

**Solution :**

| D | D | D | $(R_1 - R_2)$ | $(R_2 - R_3)$ | $(R_1 - R_3)$ | $D^2$ | $D^2$ | $D^2$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 6 | -2 | -3 | -5 | 4 | 9 | 25 |
| 6 | 5 | 4 | 1 | 1 | 2 | 1 | 1 | 4 |
| 5 | 8 | 9 | -3 | -1 | -4 | 9 | 1 | 16 |
| 10 | 4 | 8 | 6 | -4 | 2 | 36 | 16 | 4 |
| 3 | 7 | 1 | -4 | 6 | 2 | 16 | 36 | 4 |
| 2 | 10 | 2 | -8 | 8 | 0 | 64 | 64 | 0 |
| 4 | 2 | 3 | 2 | -1 | 1 | 4 | 1 | 1 |
| 9 | 1 | 10 | 8 | -9 | -1 | 64 | 81 | 1 |
| 7 | 6 | 5 | 1 | 1 | 2 | 1 | 1 | 4 |
| 8 | 9 | 7 | -1 | 2 | 1 | 1 | 4 | 1 |
| | | | | | | $\underline{200}$ | $\underline{214}$ | $\underline{60}$ |

Spearman's Coefficient of Rank Correlation

$$\gamma_s = 1 - \frac{6 \, \Sigma D^2}{N \, (N^2 - 1)}$$

Between 1 and 2     $= 1 - \dfrac{6 \times 200}{10(10^2 - 1)}$

$$= 1 - \frac{1200}{990}$$

$= 1 - 1.212$

$= -0.212$

Between 2 and 3
$$= 1 - \frac{6 \times 214}{10(10^2 - 1)}$$

$$= 1 - \frac{1284}{990}$$

$= 1 - 1.296$

$= -0.296$

Between 1 and 3
$$= 1 - \frac{6 \times 60}{10(10^2 - 1)}$$

$$= 1 - \frac{360}{990}$$

$= 1 - 0.3637$

$= 0.6363$

Since the correlation between the judges 1&3 is positive value, it can be said that the pair1st and 3rd judges have the nearest approach to common taste in beauty.

## When Ranks are repeated :

Spearman's Rank Correlation $= 1 - \dfrac{6 \left[ \in D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + - - - - -n \right]}{N(N^2 - 1)}$

## Example 4 :

Eight students have obtained the following marks in Accountancy and Economics. Calculatethe rank coefficient of correlation.

| Accountancy (x) : | 25 | 30 | 38 | 22 | 50 | 70 | 30 | 90 |
|---|---|---|---|---|---|---|---|---|

| Economics (y) : | 50 | 40 | 60 | 40 | 30 | 20 | 40 | 70 |
|---|---|---|---|---|---|---|---|---|

**Solution :**

| X | $R_1$ | y | $R_2$ | Ɖ ($R_2 - R$) | $D^2$ |
|---|---|---|---|---|---|
| 25 | 2 | 50 | 6 | -4 | 16.00 |
| 30 | 3.5 | 40 | 4 | -0.5 | 0.25 |
| 38 | 5 | 60 | 7 | -2 | 4.00 |
| 22 | 1 | 40 | 4 | -3 | 9.00 |
| 50 | 6 | 30 | 2 | +4 | 16.00 |
| 70 | 7 | 20 | 1 | +6 | 36.00 |
| 30 | 3.5 | 40 | 4 | -0.5 | 0.25 |
| 90 | 8 | 70 | 8 | 0 | 0.00 |

$\overline{N=8}$ $\qquad$ $\overline{N=8}$ $\qquad$ $\in \overline{D^2} = \underline{81.5}$

$$R_s = 1 - \frac{6 \left[\in D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + - - - - -n\right]}{N(N^2 - 1)}$$

Hence 30 is repeated twice in x series so m = 2
Hence 40 is repeated thrice in y series so m = 3

$$= 1 - \frac{6 \left[81.5 + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(3^3 - 3)\right]}{8(8^2 - 1)}$$

$$= 1 - \frac{6 \left[81.5 + 0.5 + 2\right]}{504}$$

$$= 1 - \frac{6 \left[84\right]}{504}$$

$$= 1 - \frac{504}{504}$$

## 14.4 MERITS OF RANK CORRELATION METHOD :

1. It is easy to calculate and understand as compared to Karl Pearson's coefficient ofcorrelation.

2. When the ranks of different values of the variables are given, it is then the only method left to calculate the degree of correlation.

3. When actual values are given and we are interested in using this formula, then wehave to give ranks to calculate correlation.

4. This method is employed usefully when the data is given in a qualitative nature likebeauty, honesty, intelligence etc.

## 14.5 DEMERITS OF RANK CORRELATION METHOD :

- This method cannot be employed in a grouped frequency distribution.

- If the items exceed 30, it is then difficult to find out ranks and their differences.

- This method lacks precision as compared to pearson's coefficient of correlation as all the information concerning the variables is not used. It is just possible that the difference between Rank correlation and coefficient of correlation may be very insignificant.

## 14.6 SUMMARY :

Spearman's Rank correlation is based on the ranking of different items in the variable. This method is useful where actual item values are not given, simply their ranks in the series are known. Thus it is a good measure in cases where abstract quantity of one group is correlated withthat of the other group.

## 14.7 QUESTIONS :

i. What is meant by Rank correlation ?

ii. Write down spearman's formula for rank correlation co-efficient.

iii. What are the Merits of Rank Correlation ?

iv. What are the Limitations of Rank Correlation ?

## 14.8 EXERCISE :

1. In a beauty competittion two judges ranked 12 participants as follows.

| Judge A : | 3 | 4 | 1 | 5 | 2 | 10 | 6 | 9 | 8 | 7 | 12 | 11 |
|-----------|---|---|---|---|---|----|---|---|---|---|----|----|
| Judge B : | 6 | 10 | 12 | 3 | 9 | 2 | 5 | 8 | 7 | 4 | 1 | 11 |

2. Two ladies were asked to rank seven different brands of lipsticks as listed below.

| Brands : | A | B | C | D | E | F | G |
|----------|---|---|---|---|---|---|---|
| Lady 1 : | 1 | 3 | 2 | 7 | 6 | 4 | 5 |
| Lady 2 : | 2 | 1 | 4 | 6 | 7 | 3 | 5 |

3. Ten participants in a beauty contest were ranked by three judges in the following order.

| Judge 1 : | 8 | 1 | 2 | 10 | 3 | 7 | 5 | 9 | 4 | 6 |
|-----------|---|---|---|----|---|---|---|---|---|---|
| Judge 2 : | 4 | 7 | 10 | 1 | 2 | 9 | 6 | 8 | 5 | 3 |
| Judge 3 : | 10 | 3 | 2 | 9 | 4 | 8 | 7 | 5 | 6 | 1 |

which of the 2 judges are agreeing with each other and who are against each other ?

4. In a contest, two judges ranked eight candidates in order of their performance as follows.

| Judge 1 : | 5 | 2 | 8 | 1 | 4 | 6 | 3 | 7 |
|-----------|---|---|---|---|---|---|---|---|
| Judge 2 : | 4 | 5 | 7 | 3 | 2 | 8 | 1 | 6 |

Find out the Rank Correlation.

5. Calculate the rank correlation coefficient for the following data.

| x : | 60 | 34 | 40 | 50 | 45 | 41 | 22 | 43 | 42 | 66 | 64 | 46 |
|-----|----|----|----|----|----|----|----|----|----|----|----|----|
| y : | 75 | 32 | 35 | 40 | 45 | 33 | 12 | 30 | 36 | 72 | 41 | 57 |

6. From the marks obtained by 8 students in Accountancy and statistics, compute coefficient of correlation by rank difference method.

Marks in

| Accountancy : | 60 | 15 | 20 | 28 | 12 | 40 | 80 | 20 |
|---------------|----|----|----|----|----|----|----|----|
| Statistics : | 10 | 40 | 30 | 50 | 35 | 20 | 60 | 38 |

7. Ten competitors in a voice contest are ranked by three judges in the following order.

| Judge 1 : | 1 | 6 | 5 | 10 | 3 | 2 | 4 | 9 | 7 | 8 |
|-----------|---|---|---|----|---|---|---|---|---|---|
| Judge 2 : | 3 | 5 | 8 | 4 | 7 | 10 | 2 | 1 | 6 | 9 |
| Judge 3 : | 6 | 4 | 9 | 8 | 1 | 2 | 3 | 10 | 5 | 7 |

Which 2 judges have the nearest approach to common likings invoice ? Which 2

judges have the opposite approach to common likings invoice ?

8. Find out spearman's rank correlation ?

| x : | 5 | 2 | 8 | 1 | 4 | 6 | 3 | 7 |
|-----|---|---|---|---|---|---|---|---|
| y : | 4 | 5 | 7 | 3 | 2 | 8 | 1 | 6 |

9. Eight students have obtained the following marks in accountancy and statistics. Find out rank correlation.

| x : | 56 | 48 | 40 | 67 | 75 | 80 | 85 | 35 |
|-----|----|----|----|----|----|----|----|----|
| y : | 75 | 43 | 56 | 94 | 71 | 92 | 76 | 54 |

10.    Ten participants in a beauty contest were ranked by three judges in the following order.

        Judge 1 :    8    1    2    10    3    7    5    9    4    6

        Judge 2 :    4    7    1    1    2    9    6    8    5    3

        Judge 3 :    10    3    2    9    4    8    7    5    6    1

Using rank correlation, determine which pair of judges have the nearest approach tocommon tastes in beauty.

11.    Calculate the rank coefficient of correlation from the following data.

        x :    80    78    75    75    68    67    60    59

        y :    12    13    14    14    14    16    15    17

12.  Calculate the rank coefficient of correlation from the data given below .x  :

    91                97    102 103    103    105    110    114    116 124

y :                102    94    105    115    113    99    92    112    120 108

13.    Find out Rank Correlation

        x :    10    12    60    60    70

        y :    15    20    20    20    50

**T. Nageswara Rao**
**Vice - Principal**
**Hindu College, Guntur.**