DIRECTORATE OF DISTANCE EDUCATION
UNIVERSITY OF JAMMU
JAMMU

SELF LEARNING MATERIAL
B.A. SEMESTER – I

SUBJECT : STATISTICS **UNIT : I-V**

COURSE NO. : ST - 101 **LESSON NO.: 1-23**

COURSE CO-ORDINATOR
Mr. Stanzin Shakya

Proof Reading by :
Dr. Vikas Sharma

http:/www.distanceeducationju.in

Printed and Published on behalf of the Directorate of Distance Education, University of Jammu, Jammu by the Director, DDE, University of Jammu, Jammu

DESCRIPTIVE STATISTICS AND PROBABALITY THEORY

Printed by S K Printers/ 2020 / 100

DETAILED SYLLABUS OF B.A. SEMESTER - I STATISTICS
PAPER TITLE : DESCRIPTIVE STATISTICS AND PROBABILITY TFEORY

Objectives: The objectives of this course is to impart students the basic knowledge of measures of central tendencies and measures of dispersion along with the introduction to concept of probability and its basic theory.

UNIT-I

Definition, Scope & importance of statistics, General nature of statistical data, qualitative & quantitative data, discrete & continuous data, Primary & Secondary data, classification & Tabulation, frequency distribution & their graphical& diagrammatic representations Histogram, frequency Curves, Bar diagram, Ogive & measure of central tendency (A.M.,G.M., H.M.) Median & mode, their merits& demerits.

UNIT- II

Measures of Dispersion : Range, Inter-Quartile range, Mean Deviation,Standard Deviation, Variance & Coefficient of Variation, Partition values, Moments (raw & central moments up to order four. Effect of change of origin & scale on moment. Shepherd's correction (without proof). Skewness & Kurtosis meaning & measures.

UNIT-III

Bivarite data: Scatter Diagram, product moment correlation coefficient, its properties & simple illustrations. Spearman's rank correlation coefficient, Intra class corelation Coefficient & correlation ratio. Coefficient of deterimination.

UNIT-IV

Probability:Random experiment, events, algebra of events, sample space, definitions of probability, simple illustrations for three events, conditional Probability, theorem on Probability of two events & its extension, independent events, simple illustrations, Bayes theorem & its applications.

UNIT-V

Probability mass function & Probability density function, joint marginal & conditional pmf & pdf , Jacobian Transformation for one and two variables, Independence of random variables, Discrete & continuous random variables, Mathematical expectation, expectation of sum of two random variables & production of two independent random variables, conditional expectation & conditional variance, moment generating function & properties of mgf.

NOTE FOR PAPER SETTING

The question paper will contain two sections. Section A will contain compulsory ten very short answer type questions of 1 mark each. Section B will contain 7 short answer type question of 5 marks each, atleast one questions from each unit and the student has to attempt any five questions. Section C will contain 10 long answer type questions, two from each unit, of 9 marks each and the student has to attempt five questions selecting one from each unit.

BOOKS RECOMMENDED

1. **Gupta & Kapoor : Fundamentals of Mathematical Statistics.**
2. **Kpoor & Saxena : Mathematical Statistics.**
3. **Goon, Gupta & Dass Gupta : Fundamentals of Statistics Vol-I**
4. **S.P. Gupta : Statistical Methods.**
5. **Croxton F.E., Cowden D.J. & Kelin S : Applied General Statistics, Prentice Hall of India.**
6. **Mood A.M., Boes D.C. & Grybill F.A. : Introduction to the Theory of Statistics.**
7. **Parazen : Modern Prob. Theory.**
8. **M.N. Murthy : Introduction to the theory of probability.**
9. **V.K. Rohatgi : Introduction to the theory of Probability.**

TITLE OF THE COURSE

CONTENTS

UNIT - I **LESSON - 1**

DEFINITIONS, SCOPE AND IMPORTANCE OF STATISTICS

STRUCTURE:

1.1 INTRODUCTION:-

The word *statistics*, when referring to the scientific discipline, is singular, as in "Statistics is an art. This should not be confused with the word *statistic*, referring to a quantity calculated from a set of data, whose plural is *statistics.*

Some scholars pinpoint the origin of statistics to 1663, with the publication of *Natural and Political Observations upon the Bills of Mortality* by Jhon .Early applications of statistical thinking revolved around the needs of states to base policy on demographic and economic data. The scope of the discipline of statistics broadened in the early 19th century to include the collection and analysis of data in general. Today, statistics is widely employed in government, business, and the natural and social sciences.

Its mathematical foundations were laid in the 17th century with the development of probability theory by Blaise pascal and Pierre de fermat. Probability theory arose from the study of games of chance. The method of least squares was first described by Carl Friedrich Gauss around 1794. The use of modern computers has expedited large-scale statistical computation, and has also made possible new methods that are impractical to perform manually.

1.2 OBJECTIVE:

Almost every aspect of natural phenomena and human and other activity is now subjected to measurement and interpretation in terms of statistics. The figures have become so important in human affairs that they form the basis of rational thinking and rational decisions in many spheres of human activities and events are proving that the decisions based on figures field better results. The reason being that the knowledge obtained with the help of satisfied methods and supported by numerical facts is accurate and precise. Through application of appropriate statistical methods, current performance may be measured, significant relationships may be studied, past experience may be analyzed and probable future trends appraised. To acquire knowledge and to express it precisely, statistical methods and statistics, have a vital role to play

1.3 ORIGIN AND DEVELOPMENT OF STATISTICS:

The term 'STATISTICS' has been derived from the Latin word 'STATUS' which means political state. More directly, the Italian word 'STATUS' seems to be the origin of the term 'statistics'. The term was used in the 15^{th} century for state. Germans have used this word in the same sense and they spelled it as 'STATISTIK'. These words mean 'political state' or the 'statesman's art'. In 1770, Baron J.F.Von Bielfied defined statistics in his book. 'The elements of universal Erudition', as "The science that teaches what is the political arrangement of all the modern states of the known world".

In Germany, systematic collection of statistics by the state started during the end of the 18^{th} century. In England it was started during Napoleonic wars because to raise revenue to meet the wars expenses it was thought proper to bring precisions in the statistics of public revenue and expenditure.

In India, too, the collection of Statistics is an old age tradition. In the ancient works like 'Manusmirity',' Shukraniti' etc, there is description of methods of organization for the collection of Statistics for running the state.

Magasthenese had given an account of method of collection of data regarding revenue and expenditure, births and deaths, military, land etc, during Chandragupta's time. Kautalya's Arthashatra too describes all these. During Mughal period, Statistics used to be collected and the system of collection was described in 'Tuzuk-i-Babri' and ' Ain-i-Akbari'. During Akbar's time, Raja Toder mal collected land Statistics for determining revenue.

In India, Prof. P.C. Mahalonobis , has contributed a lot in the theoretical and applied field of Statistics .In the applied field the names of C.R.Rao, R.C. Desai, Dr. P.V.Sukhatme etc are also noteworthy.

Statistics has now become a universal applicable science. Statistics are required as a basis of action in all fields whenever information can be measured .

1.4 DEFINITION OF STATISTICS:

Statistics has been defined differently by different authors from time to time. To a layman, it simply means a mass of figures of collection of data. It is true that Statistics deals with aggregates of figures, but all sorts of numerical data are not Statistics and before we attempt a formal definition of Statistics, it is necessary to give some explanations. In physical sciences, experimental methods are used to know the effect of a particular cause.

Statistics is a science of collecting, analysis and interpretating the numerical data. The word 'Statistics 'is used in singular and plural senses.

The most exhaustive definition has been given by Prof. Horace Secrist. He defines:

" By Statistics we mean aggregate of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standard of accuracy, collected in a systematic manner for a pre-determined purpose and placed in the relation to each other".

This definition mentions the characteristics which data, the subject- matter of

Statistics, should posses.

According to Dr. A.L.Bowley "Statistics are numerical statements of facts in any department of enquiry placed in relation to each other."

Sir R.A.Fisher defined Statistics as "The science of Statistics is essentially a branch of applied mathematics and may be regarded as mathematics applied to observational studies".

According to Lovitt: Statistics is science which deals with the collecting, classifying, presenting, comparing and interpreting numerical data collected to throw the light on any sphere of enquiry".

The definition of Statistics, given according to modern concepts, i) Make the scope of the subject comprehensive, and ii) include all statistical methods. Some of the definitions are given below:

According to W.I.King, "The science of Statistics is the method of judging collective natural or social phenomena from the results obtained by the analysis of an enumeration or collection of estimates."

According to the definition,

a) natural or social phenomena are studied in Statistics

b) the basis of study is the data collected either by enumeration or estimation , and

c) The study is by nature collective, individual facts are not studied.

In the words of Prof. Lovitt," Statistics deals with the collection, classification and tabulation of numerical facts as the basis for explanation, description and comparison of phenomena." Prof. Lovitt's definition has two parts, first, explanation, description and comparison of phenomena are the objects of Statistics and second collection, classification and tabulation are the means to achieve these objectives.

According to F.E.Croxton and D.J.Cowden," Statistics or Statistical methods may be defined as the collection, presentation, analysis and interpretation of numerical data Ya-Lun-Chou defines Statistics as "Statistics is the science that deals with the analysis of masses of quantitative data. It includes the collection, classification, presentation and interpretation of such data."

1.5 SCOPE OF STATISTICS:

Almost every aspect of natural phenomena and human and other activity is now subjected to measurement and interpretation in terms of statistics. The figures have become so important in human affairs that they form the basis of rational thinking and rational decisions in many spheres of human activities and events are proving that the decisions based on figures field better results. The reason being that the knowledge obtained with the help of satisfied methods and supported by numerical facts is accurate and precise. Through application of appropriate statistical methods, current performance may be measured, significant relationships may be studied, past experience may be analyzed and probable future trends appraised. To acquire knowledge and to express it precisely, statistical methods and statistics, have a vital role to play.

Statisticians improve the quality of data with the design of experiments and surveys sampling. Statistics also provides tools for rediction and forecasting using data and Statistical models. Statistics is applicable to a wide variety of academic disciplines, including natural and social science, government, and business.

Statistical methods can be used to summarize or describe a collection of data; this is called descriptive statistics. This is useful in research, when communicating the results of experiments. In addition, patterns in the data may be modelled in a way that accounts for randomness and uncertainty in the observations, and are then used to draw inferences about the process or population being studied; this is called inferential statistics. Inference is a vital element of scientific advance, since it provides a prediction (based in data) for where a theory logically leads. To further prove the guiding theory, these predictions are tested as well, as part of the scientific method. If the inference holds true, then the descriptive statistics of the new data increase the soundness of that hypothesis.

1.6 IMPORTANCE OF STATISTICS:

Statistics plays a vital role in every fields of human activity. Statistics has important role in determining the existing position of per capita income, unemployment, population growth rate, housing, schooling medical facilities etc…in a country. Now statistics holds a central position in almost every field like Industry, Commerce, Trade, Physics, Chemistry, Economics, Mathematics, Biology, Botany, Psychology, Astronomy etc…, so application of statistics is very wide. Now we discuss some important fields in which statistics is commonly applied.

(1) <u>In Business</u>

Statistics play an important role in business. A successful businessman must be very quick and accurate in decision making. He knows that what his customers wants, he should therefore, know what to produce and sell and in what quantities. Statistics helps businessman to plan production according to the taste of the costumers, the quality of the products can also be checked more efficiently by using statistical methods. So all the activities of the businessman based on statistical information. He can make correct decision about the location of business, marketing of the products, financial resources etc…

(2) <u>In Economics</u>:

Statistics play an important role in economics. Economics largely depends upon statistics. National income accounts are multipurpose indicators for the economists and administrators. Statistical methods are used for preparation of these accounts. In economics research statistical methods are used for collecting and analysis the data and testing hypothesis. The relationship between supply and demands is studies by statistical methods, the imports and exports, the inflation rate, the per capita income are the problems which require good knowledge of statistics.

(3) <u>In Mathematics</u>:

Statistical plays a central role in almost all natural and social sciences. The methods of natural sciences are most reliable but conclusions draw from them are only probable, because they are based on incomplete evidence. Statistical helps in describing these measurements more precisely. Statistics is branch of applied mathematics. The large number of statistical methods like probability averages, dispersions, estimation etc… is used in mathematics and different techniques of pure mathematics like integration, differentiation and algebra are used in statistics.

(4) <u>In Banking</u>:

Statistics play an important role in banking. The banks make use of statistics for a number of purposes. The banks work on the principle that all the people who deposit their money with the banks do not withdraw it at the same time. The bank earns profits out of these deposits by lending to others on interest. The bankers use statistical approaches based on probability to estimate the numbers of depositors and their claims for a certain day.

(5) In State Management (Administration):

Statistics is essential for a country. Different policies of the government are based on statistics. Statistical data are now widely used in taking all administrative decisions. Suppose if the government wants to revise the pay scales of employees in view of an increase in the living cost, statistical methods will be used to determine the rise in the cost of living. Preparation of federal and provincial government budgets mainly depends upon statistics because it helps in estimating the expected expenditures and revenue from different sources. So statistics are the eyes of administration of the state.

(6) In Accounting and Auditing:

Accounting is impossible without exactness. But for decision making purpose, so much precision is not essential the decision may be taken on the basis of approximation, know as statistics. The correction of the values of current asserts is made on the basis of the purchasing power of money or the current value of it.

In auditing sampling techniques are commonly used. An auditor determines the sample size of the book to be audited on the basis of error.

(7) In Natural and Social Sciences:

Statistics plays a vital role in almost all the natural and social sciences. Statistical methods are commonly used for analyzing the experiments results, testing their significance in Biology, Physics, Chemistry, Mathematics, Meteorology, Research chambers of commerce, Sociology, Business, Public Administration, Communication and Information Technology etc…

(8) In Astronomy:

Astronomy is one of the oldest branch of statistical study, it deals with the measurement of distance, sizes, masses and densities of heavenly bodies by means of observations. During these measurements errors are unavoidable so most probable measurements are founded by using statistical methods.

Example: This distance of moon from the earth is measured. Since old days the astronomers have been statistical methods like method of least squares for finding the movements of stars.

Statistics in War:

In war, the theory of decision functions can be of great assistance to the military and technical personnel to plan 'maximum destruction with minimum efforts'. Thus, we see that that the science of Statistics is associated with almost all the sciences social as well as physical.

1.7 LIMITATIONS OF STATSTICS

Does not deal with individual measurement.

Deals only with quantities characteristics.

Result is true only on an average.

It is only one of the methods of studying a problem.

Statistics can be measured. It requires skills to use it effectively, otherwise misinterpretation is possible

It is only a tool or means to an end and not the end itself which has to be intelligently identified using this tool.

1.8 DISCRETE AND CONTINUOUS DATA:

A type of data is **discrete** if there are only a finite number of values possible or if there is a space on the number line between each 2 possible values.

Ex. A 5 question quiz is given in a Math class. The number of correct answers on a student's quiz is an example of discrete data. The number of correct answers would have to be one of the following : 0, 1, 2, 3, 4, or 5. There are not an infinite number of values, therefore this data is discrete. Also, if we were to draw a number line and place each possible value on it, we would see a space between each pair of values.

Discrete data usually occurs in a case where there are only a certain number of values, or when we are counting something (using whole numbers).

Continuous data makes up the rest of numerical data. This is a type of data that is usually associated with some sort of physical measurement.

One general way to tell if data is continuous is to ask yourself if it is possible for the data to take on values that are fractions or decimals. If your answer is yes, this is usually continuous data.

Ex. The length of time it takes for a light bulb to burn out is an example of continuous data. Could it take 800 hours? How about 800.7? 800.7354? The answer to all 3 is yes.

Illustration:

1) Discrete data (ungrouped):

No. of days	No. of accident
0	12
1	9
2	7
3	6
4	4
5	3
6	4

2) Discrete data (grouped):

No. of deaths	Days
2-3	5
3-4	3
4-5	2
5-6	1

1) Continuous data :

Heights of the students (in cms)	No. of students
50-100	7
100-150	11
150-200	8

1.9 PRIMARY DATA AND SECONDARY DATA:

Primary data are obtained by a study specially designed to fulfill the data needs of the problem in hand. Such data are original in character and are generated in large number of surveys conducted mostly by the Government and also by some individuals, institutions and research bodies. For example, data obtained in a population census by the office of the Registrar General and census Commissioner, ministry of home affairs, are primary data.

Data which are not originally collected but rather obtained from published or unpublished sources are known as secondary data . For example, the office of the Registrar General and Census commissioner the census data are primary whereas for all others, who use such data, that is secondary . The secondary data constitute the chief material on the basis of which Statistical work is carried out in many investigations. In fact, before collecting primary data it is desirable that one should go through the existing literature and learn what is already known of the general area in which the specific problem falls and any all surrounding information that may give us leads the lessons. This can help in getting an idea about the possible pitfalls, avoiding duplication of efforts and waste of resources . It should be noted that it is process of 'compiling' Statistics (i.e. secondary data) from various published sources. To quote Crum, Patton and Tabbutt,' collection means the assembling, for the purpose of a particular investigation of entirely new data, presumably not already available in published sources'. We have used the 'collection' in the book strictly in the narrow sense defined above.

The difference between Primary and Secondary data is only the degree – data which are primary in the hands of one become secondary in the hands of another. Data are primary for the individual agency or institution collecting them whereas for the rest of the world they are secondary . A few examples would clarify the distinction. Suppose, an Investigator wants data about the spending habits of the students of Delhi university. If he collects the data himself or through his agents adopting any suitable method such as contacting and interviewing students or circulating a questionnaire, the data would constitute primary data for him. On the other hand, if the students union has already made similar survey and the investigator obtained data from the union office, such data would constitute secondary data for him. Similarly, Statistics collected by various departments of the government such as Labour Bureau and the central Statistical Organization are primary for the respective departments whereas for all other they constitute secondary data.

Secondary data offers the following advantages:

1) It is highly convenient to use information which someone else has compiled. There is no need for printing data collection forms, vhiring enumerators, editing data and tabulating the results, etc. Researchers alone or with some clerical assistance may obtain information from published records complied by some body else.

2) If secondary data are available on some subjects they are more quicker to obtain than primary data.

3) Secondary data may be available on some subjects where it would be impossible to collect primary data. For example, census data cannot be collected by an individual or research organization, but can only be obtained from Government publications.

Sources of Secondary data :

In most of the studies the investigator finds it impractical to collect first hand data on all the related issuses and as such he makes use of the data collected by others . There is a vast amount of published information from Statistics studies may be made and fresh Statistics are constantly in the state of production.The sources of secondary data are :

1) Published sources.

The various sources of published data are :

i) Reports and official publications of -

a) Internationally bodies such as the 'World Bank', ' International Labour Organisation', 'Statistical Office of the United nation'.

b) Reports of the Adhoc committees and commissions appointed by the government such as Scapital Committee, Mehrotra Committee, Shah commission etc.

c) Central and state Governments such as Abstract of the Indian union, Economic survey of India, India1988-89, etc.

ii) Semi- official publications of various local bodies such as Municipal corporation and District Boards.

iii) Publications of autonomous and private institutes such as :

a) Trade and professional bodies, such as , the federation of Indian chambers of commerce and industry , the Institute of charted Accountants, the institute of foreign trade, The prestigious journals of these institutes are respectively " Economic trends" , 'The charted Accountant' , 'The Foreign Trade'.

b) Financial and Economic journals such as ' Indian Economic Review' , Reserve Bank of India Bulletin' , Indian Finance' .

c) Annual reports of joint Stock companies and corporations.

d) Publications brought out by various autonomous research Institutes and Scholars such as Institute Economic Growth, Delhi; National council of Applied Economic research, New Delhi; Institute of Politics and Economics, Pune.

It should be noted that the publications mentioned above vary with regard to the periodicity of publication. Some are published at regular intervals (yearly, monthly, weekly, etc.,) whereas other are adhoc publications, i.e., with no regularity about periodicity of publication.

2) **Unpublished sources :**

All Statistical material is not published. There are various sources of unpublished data such as records maintained by various Government and Private offices, studies made by research institutions, Scholars, etc. Such sources can be used where necessary.

Some Government Publications:

1) Statistical Abstract- Annual, CSO
2) Monthly Abstact of Statistics –India, CSO.
3) Savekshana, NSSO(Quarterly).
4) Census of India reports, RGI
5) Sample Registration Bulletin (Quarterly), RGI
6) Monthly Statistics of Foreign Trade of India, DGCIS
7) Indian Forest Statistics (Annual), DESCRIBED

8) Agricultral situation in India (monthly), DESCRIBED

9) Indian Labour Statistics, Indian Labour journal.

10) Monthly Index of industrial Production, CSO , etc.

1.10 ASSESSMENT QUESTIONS :

1. Define Statistics. Gives its scope and importance.
2. Distinguish between discrete and continuous data.
3. Distinguish between primary and secondary data.
4. What are various sources of secondary data.

****

UNIT -1 **LESSON - 2**

CLASSIFICATION AND TABULATION

STRUCTURE:

2.1 INTRODUCTION:

The collected data is usually contained in schedules and questionnaires. But that is not in an easily assailable form. The answers will require some analysis if their salient points are to be brought out. As a rule, the first step in the analysis is to classify an tabulate the information collected, or if published statistics have been employed rearrange these into new groups an tabulate the new arrangement. In case of some investigations, the classification and tabulation may give such a clear picture of the significance of the material arranged that no further analysis is required. They are however very important whether they complete the analysis or form only part of it. The questionnaire may have been very carefully drawn up and the whole information displayed in tabular form, no one will be a great deal wiser as to the contents of the replies.

2.2 OBJECTIVES OF CLASSIFICATION:

1. To condense the mass of data:
2. To enable grasping of data:
3. To prepare the data for tabulation

2.3 MEANING OF CLASSIFICATION:

Classification is a process of arranging data into different classes according to their resemblances and affinities. The arrangement of a huge mass of heterogeneous data into homogeneous groups facilitates comparison and analysis of the data. Classification prepares the ground for the proper presentation of statistical facts.

After collection and editing of data the first step towards further processing the same is classification. Classification is the grouping of related facts into classes. Facts in one class differ from those of classification. sorting facts on one basis of classification and then on another basis is called cross- classification. This process can be repeated as many times as there are possible bases of classification. Classification of data is a function very similar to that of sorting letters in a post- office are sorted into different lots on a geographical basis, i.e., in accordance with their destinations such as Mumbai, Calcutta, Kanpur, Jaipur, etc. They are then put into separate bags , each containing letters with a common characteristic, viz., having the same destination. To take another example, when students seek admission in a college they submit applications to the office. The applications forms contain particulars about their performance in the previous examinations, their date of birth, sex, nationality, etc. If one is interested in finding out how many first, second and third class students have joined the college, one may look into each and every form and note whether it relates to a first class student, second class student, etc. He may find that out of 1,000 students who took admission 50 had first class ,800 second class and 150 third class. The process with the help of which this information in a summary form is obtained is called the classification of data

DEFINITIONS:

"Classification is the process of arranging things (either actually or notionally) in groups or classes according to their resemblances and affinities and gives expression to the unity of attributes that may subsist amongst a diversity of individuals."

- Conner

"Classification is the process of arranging data into sequences and groups according to their common characteristics, or separating them into different but related parts

2.4 FUNCTIONS OF CLASSIFICATION:

1 Bulk of the data

2. Simplifies the data

3 Facilitates comparison of characteristics

4 Renders the data ready for statistical analysis

Tabulation of data:

After classification the next essential step is that of tabulation in which the data are put in a table having different rows and columns. The process by which the classified data are presented in an orderly manner by being placed in proper rows and columns of a table in order to bring out their chief characteristics and essential features is known as tabulation . It helps to get the answer of every type of enquiry pertaining to that data at a glance only, and makes the data ready for analysis.

Tabulation is of two types:

1) Simple tabulation.

2) Complex tabulation.

1) Simple tabulation (one way table):

In this type of tabulation , table contains information pertaining to only one set of data and gives information about one or more groups of independent questions. For example

	State	Population
1.	Uttar Pradesh	……………
2.	Madhya Pradesh	……………
3.	Bihar	……………
4.	M.P	……………
5.	A.P	……………
6.	Punjab	……………

2) Complex Tabulation :

Complex Tabulation includes the following cases:

a) Double table or two way table

b) Treble table or three way table.

b) Manifold table

a) Double table or two way table:

Here, the numerical data are classified according to two characteristics and tabulated. For example, the population of India can be divided according to the states and it can further be subdivided according to sex. Such table is known as double table or two way table.

State	Population		
	Male	female	Total
1. Uttar Pradesh	……………	……………	……………
2. Madhya Pradesh	……………	……………	……………
3. Bihar	……………..	…………...	…………..
4. J&K	……………..	…………...	…………..

b) Treble Table or three way table:

In this type of data are classified according to three characteristics and tabulated. For example, the population of India, after being divided according to states, is further sub divided according to their sex and ultimate divisions are for the last time divided according to their literacy. The different rows and columns for this type will be shown below. Such table is known as three way or treble table.

States	Population					
	Males		Females		Total	
	Edu.	Unedu.	Edu.	Unedu.	Edu.	Unedu.
1. Uttar Pradesh						
2. Madhya Pradesh						
3. Punjab						
4. Harayana						
Totals						

c) **Manifold Table:**

When the data are classified according to more than three characteristics, they are tabulated in a table known as manifold table.

Difference between Classification and Tabulation:

Classification and Tabulation are of great importance in Statistical investigations. In carrying out the process of classification, to make the study complete and accurate in every respect, it is natural to arrange the results in a tabular form. In actual practice first we classify the data According to some rule and then represent the classified data in the tabular form.

Thus classification is the basis of tabulation. In the process of classification, the raw data is divided into groups on the basis of similarity and dissimilarity, while in the tabulation the different groups are distinguished by lines of different dimensions or different colours. In this way, tabulation is the mechanical function of classification.

Thus classification and tabulation are not two distinct processes. However, a distinction is often made between classification and tabulation.

Classification	Tabulation
1. First the data are classified. Classification is not based on tabulation.	1. The classified data is tabulated. Thus tabulation is not based on classification.
2. The data is divided into different groups on the basis of similarity and dissimilarity	2. The data is distinguished by lines of different dimensions or colours.

3. Classification is the process of Statistical analysis	3. Tabulation is the method of presentation of data.
4. Here the data are divided into groups or Subgroup	4. Here the data are presented in the form of headings and sub-headings.
5. Classification is not a mechanical function of tabulation.	5. Tabulation is a mechanical function of classification.

Examples:

1) In 1985, out of a total 4000 workers of a factory, 2500 were permanent. The number of women workers was 500 out of which 300 were temporary. In 1986, the number of workers increased to 3000 of which 2000 were men. On the other hand the numbers of temporary workers fall down to 250 of which 150 were women. Present the above data in the form of a appropriate table.

2) Tabulate the following information:

   In a trip organized by a college there were 100 persons, each of whom paid Rs. 25.50 on an average. There were 70 students each of whom paid Rs. 26. Members of the teaching staff were charged nothing. The number of ladies was 20% of the total of which one was a lady staff member.

2.5 DIAGRAMMATIC AND GRAPHICAL REPRESENTATION OF THE DATA:

'The representation of quantitative data suitably through charts and diagrams is known as Graphical Representation of Statistical Information. Graphs include both charts and diagrams. There are various types of graphs in the form of charts and diagrams. Some of them are

(a) Line Chart

(b) Bar Chart

(c) Multiple bar chart

(d) Sub divided bar chart

(e) Pie Chart

a) **Line Chart:**

Line Chart are simple mathematical graphs that are drawn on the graph paper by plotting the data concerning one variable on the horizontal x-axis and other variable of data on the vertical y-axis. With the help of such graphs the effect of one variable upon another variable during an experimental or normative study may be clearly demonstrated. The construction of these graphs can be understood through the following example.

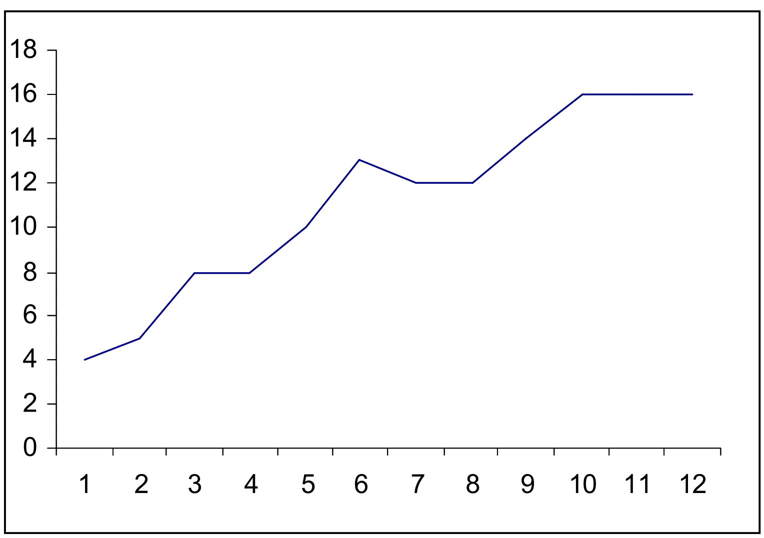
Example: The following data shows the number of accidents sustained by 134 drivers of a public utility company over a period of five years

No. of accidents:	1	2	3	4	5	6	7	8	9	10	11	12
No. of drivers:	4	5	8	8	10	13	12	12	14	16	16	16

Represents the data by line chart.

Solution: Plot the points

Line chart



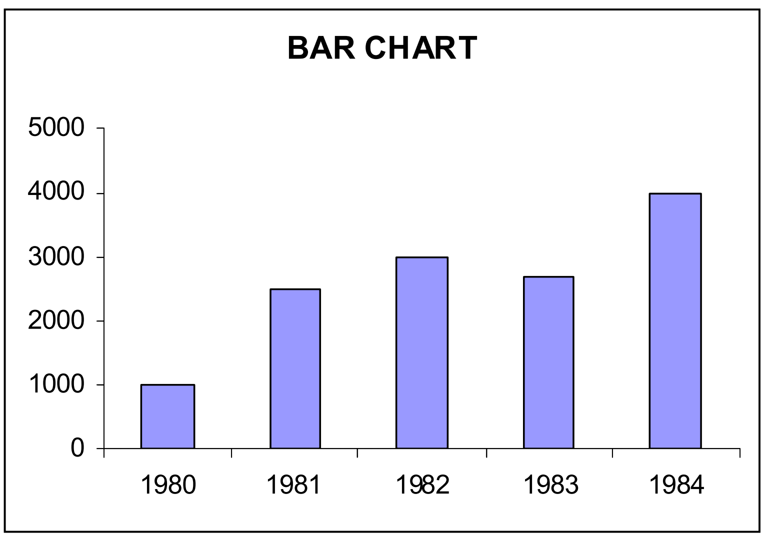
b) **Bar Chart**:

Another mode of diagrammatic representation of data is the use of bar diagrams. These have more general applicability than line diagrams in the sense that they may be used for series varying either over time or over space. In this method, bars of equal width are taken for the different items of the series, the length of a bar representing the value of the variable concerned. It is preferable to take the bars horizontally for data varying over space and vertically in the case of a series varying over time.

Example: Draw a bar chart of the procurement of rice (in tons) in an Indian State

Solution:

Year :	1980	1981	1982	1983	1984
Prod.:	1000	2500	2900	2400	3500



The given data is represented by the simple bar Diagram as only one variable to be represented.

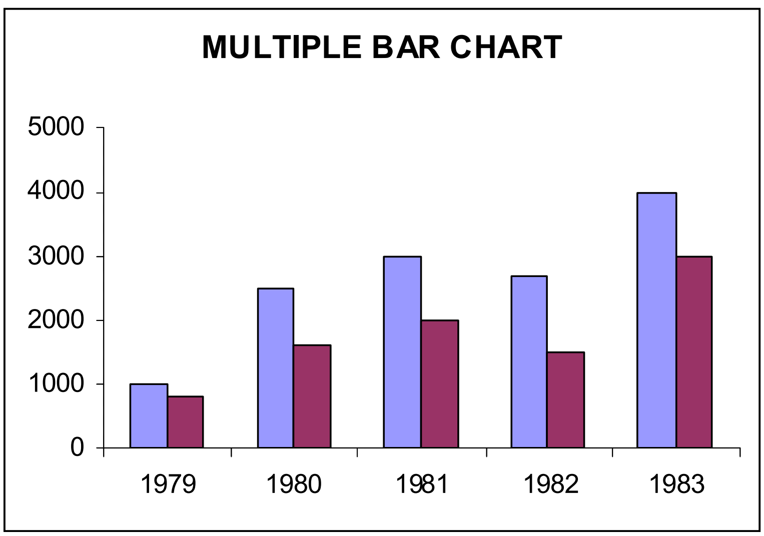
A variant of the bar diagram is the multiple bar diagram, which is employed in comparing two or more series of data on the same variable. Thus we may want to compare the population figures, as recorded in a number of censuses, for two or more countries, or we may like to compare the yield of paddy for a number of States for two or more time periods.

c) **Multiple Bar Diagram**: Multiple bar diagrams supply information about more than one phenomenon. If two or more sets of inter related or variable are to be presented graphically, multiple bar diagrams are used.

Example: The data below give the yearly profits(in thousand of rupees) of two companies A and B.

Year	Company (A)	company (B)
1979	1000	800
1980	2500	1500
1981	3000	2000
1982	2700	1200
1983	3800	3200

Represent the following data by means of suitable diagram.



(e) **Sub divided bar chart**:

The bar diagram may also be used to exhibits the division of a whole into its components parts.

Example: Represent the following data by a suitable diagram:

Items Expenditure Family income (Rs. 1000)

Food	250
Clothing	225
Education	125
Miscellaneous	290
Deficit	110

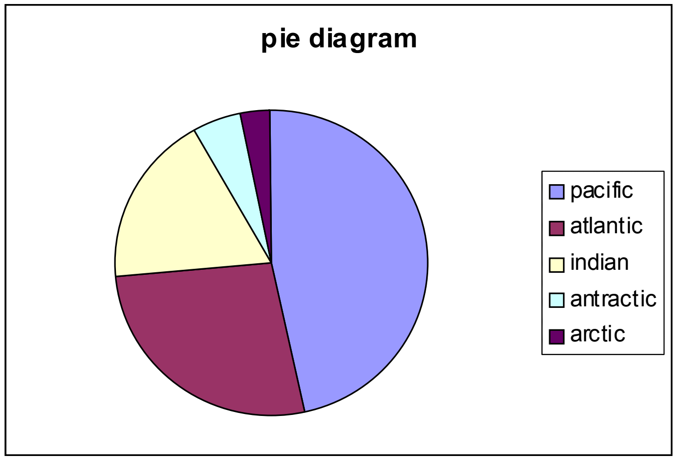
Solution: The data can be represented by sub-divided bar diagram (do it your self)

f) **Pie Diagram**:

The pie chart is a circular graph which represents the total value with its components. The area of a circle represents the total value and the different sectors of the circle represent the different parts. The circle is divided into sectors by radii and the areas of the sectors are proportional to the angles and centre. It is generally used for comparing the relation between various components of a value and between components and the total value. In pie chart, the data is expressed as percentage. Each component is expressed as percentage of the total value. The pie chart is also known as circular chart or sector chart.The name pie diagram is given to a circle diagram because in determining the circumference of a circle we have to take into consideration a quantity known as pie (written as).The surface area of a circle is known to cover 2radians or 360 degrees. The data to be represented through a circle diagram may therefore be presented through 360 degrees, parts or sections of a circle. The total frequencies or value is equated to 360° and then the angles corresponding to component parts are calculated. After determining these angles, the required sectors in the circle are drawn.

Example: The following table shows the area in millions of sq.km., of oceans of wo

Oceans	Area(million sq.km)
Pacific	70.8
Atlantic	41.2
Indian	28.5
Antractic	7.6
Arctic	4.8



2.6 SELF ASSESSMENT QUESTIONS:

1. Discuss the usefulness of diagrammatic representation of facts.
2. Briefly explain the principles of classification.
3. In a recent study of strikes in mills, an experimenter collected the following data.

Causes :	economic	personal	political	rivalry	others
(in percentage): Occurrences	58	16	10	6	10

Represent the data by bar chart.

4. A Rupee spent on "Khadi" is distributed as follows:s

	paise
Farmer	19
Carder and spinner	35
Weaver	28
Washerman, dyer and printer	8
Administrative agency	10
Total:	100

Present the data in the form of a pie diagram.

5. What is tabulation? What are its uses? Mention the items that a good statistical table should contain.

****

UNIT - 1 **LESSON - 3**

FREQUENCY DISTRIBUTION

STRUCTURE:

3.1 INTRODUCTION:

In the previous lesson, we have discussed some elementary methods of dealing numerical form in which they are collected. In this lesson, we shall discuss in the detailed form of frequency tables. However, before we set out to discuss the construction of frequency tables, it may be helpful to explain some important related concepts which are discussed as definitions.

3.2 OBJECTIVES:

After reading this lesson, you should be able to :

· Define some basic definitions.

· Draw up a frequency table.

· Know about the Histogram, Frequency Polygon and Ogives.

· Know about the graphical determination of median, Quartiles and mode.

3.3 DEFINITIONS:

a) **Frequency:** The number of times the value of a variate is repeated is called frequency of that variate value. For example, If there are 5 girls who secured 25 marks in the paper then the frequency of 25 marks is 5 . If there are are 15 families with members between 0–5 , then 15 is the frequency of group 0-5.

b) **Frequency Distribution** :

The premise of data in the form of frequency distribution describes the basic pattern which the data assumes in the mass. Frequency distribution gives a better picture of the pattern of data if the number of observations is large enough.

c) **Class limits:**

The pairs of numbers that represent the class intervals are called class limits. In a specified class interval the greatest value that a variable can take is called the upper class – limit and the least value that it can take in that class interval is called the lower class limit. The true class limits are called class boundaries and the observed limits are known as class limits or class marks.

d) **Width of a class interval** :

The distance between the upper and lower class limits of a particular class interval is called the width of that class interval.

e) **Mid value** :

It is the mid point of a class interval and is obtained by adding the lower and upper limit of a class divided by 2.

f) **Class Interval**:

The difference between the Upper and the Lower limit of a class is known as class interval of that class. For example class interval of the class 500-600 is 100(i.e. 600 minus 500). Deciding class interval is an important factor while constructing a frequency distribution. It depends upon number of factors like range of the data, the details required and number of classes to be formed etc.Class intervals are of two types:

i) **Exclusive class intervals**:

In the exclusive class intervals, the upper class limit of a class interval is the lower limit of the next class interval i.e., upper class limit is not included in that class interval

e.g.

Income(Rs.)	*No. of persons*
1000-1100	50
1100-1200	100
1200-1300	200
1300-1400	150
1400-1500	40
1500-1600	10
	Total 550

In this example, there are 150 persons who have income between 1300 and 1399.99. A person whose income is 1400 will be included in the class 1400-1500.

Exclusive method ensures the continuity of data as the upper limit of one class is lower limit of the next class.

This method is widely used in practice; however it is quite

ii) Inclusive class interval:

In the inclusive class intervals, the upper limit of a class is not the lower limit of the next class interval i.e., both the limits of a class are included in that class.e.g.

Income(Rs.)	*No. of persons*
1000-1099	50
1100-1199	100
1200-1299	200
1300-1399	150

1400-1499	40
1500-1599	10
Total	550

In the class 1000-1099 persons whose income is between Rs.1000 and Rs.1099. If the income of a person is exactly 1100 he is included in the next class. This method is not suitable for continuous variable.

a) **Cumulative frequency** :

The total frequency of all values less than or greater than the upper or lower limit of a given class is called the cumulative frequency up to and including that class and a table presenting such cumulative frequencies is called cumulative frequency distribution or table .

3.4 TYPES OF FREQUENCY DISTRIBUTION:

a**)** **Discrete frequency distribution**

Formation of a discrete frequency distribution:

It is quite a simple process. We have just to count the number of times a particular value is repeated which is called the frequency of that class. In order to make the counting process easy, prepare a column of tally marks. In another column, place all possible values of variable from the lowest to the highest. For tally marks, put a bar for every repetition of a variable. After 4th bar put a cross across 4 bars which makes total of 5. This block of 5 bars is to facilitate counting. We finally count the number of bars and get the frequency of the class.

The method shall be clear from the following example.

Example: Count the number of letters in each word of the para given below (ignoring comma, full-stop, etc) and prepare a discrete frequency distribution.

"Today, to a very striking degree, our culture has become a statistical culture. Even a person, who may never have heard of an index number, is attached in an intimate fashion by the gyrations of those index numbers which describe the cost of living."

Solution:

No. of letters	Tally marks	Frequency
1	III	3
2	~~IIII~~ IIII	9
3	~~IIII~~ I	6
4	IIII	4
5	~~IIII~~ II	7
6	~~IIII~~	5
7	IIII	4
8	IIII	4
9	I	1
10		–
11	I	1

b) **Continuous or grouped frequency distribution**

For making a grouped frequency distribution we have the following steps:

1) Determine the range of the raw data.

2) Determine the number of equal parts (classes) into which the entire range shall be divided, which depends on the size and nature of the data.

3) Arrange the sheet with three headings: Classes, Tally marks and frequency.

4) Read off the observations in the raw data and for each one record a mark into the respective classes. To facilitate counting the tally marks are groups of 5. every fifth tally being drawn across the preceding four.

5) Write the sum of the tally marks in each row in the frequency distribution.

Example :

Classifies the following data by taking class interval such that their mid-values are 17, 22, 27 and 32 and so on.

30	42	30	54	40	48	15	17	51	42	24	41
30	27	42	36	28	26	37	54	44	31	36	40
36	22	30	31	19	48	16	42	32	21	22	46
33	41	21									

Solution:

Since we have to classify the data in such a manner that the mid-values are 17, 22, 27 etc the first class should be 15-19 (mid-value = (15+19)/2 = 17), second class 20-24 and so on. Therefore we get following frequency distribution table.

Class	Tally marks	Frequency
15-19	I I I I	4
20-24	I I I I	4
25-29	I I I I	4
30-34	~~IIII~~ I I I	8
35-39	I I I I	4
40-44	~~IIII~~ I I I I	9
45-49	I I I	3
50-54	I I I	3
	Total	39

Graphical representation of frequency distribution:

It is often useful to represent a frequency distribution by means of a diagram which makes the data easy to interpret and intelligible. Diagrammatic representation also facilitates comparisons of two or more observations. A frequency distribution can be presented graphically in any of the following ways:

* Histogram
* Frequency polygon
* Smoothed frequency curve
* Ogives or Cumulative Frequency curves

3.5 HISTOGRAM:

A histogram or a frequency histogram is like a bar chart, where the horizontal axis gives the interval of observations involved, and the vertical axis gives the number of observations that fall in each interval. Thus, it gives the frequencies or the number of data points of the observations in each of the intervals. The histogram is the most common way to graphically display data. Compared to the stem-and-leaf plot, it loses some information (the exact value of each data point) but it is extremely informative about the centering, spread and in general shape of the distribution of data.

The histogram should be clearly distinguished from a bar diagram. Bar diagram is one dimensional i.e. only length of the bar is kept into consideration and not the width while histogram is two-dimensional, i.e., both the length as well as width is important.

While constructing histogram the variable is always taken on the X-axis and the frequencies depending on it on the Y-axis. Each class is then represented by a distance on the scale that is proportional to its class interval. Frequencies of each class constitute the height of the rectangle and class-interval distance represents its width. If, however, the classes are of unequal width then the height of the rectangle will be proportional to the ratio of the frequencies to the width of the classes. The diagram of continuous rectangles so formed is called histogram. The area of the histogram represents the total frequency as distributed throughout the classes.

Construction of histogram

(i) **For distributions having equal class-intervals**

Take frequency on the Y-axis, the variable on the X-axis and construct adjacent rectangles. In such a case the height of the rectangles will be proportional to the frequencies.

(ii) For distributions having unequal class intervals.

When class-intervals are unequal, a correction for unequal class intervals must be made. The correction consists of finding for each class the frequency density or the relative frequency density. The frequency density is the frequency for that class divided by the width of that class. A histogram or frequency density polygon constructed from these density values would have the same general appearance as the corresponding graphical display developed from equal class intervals.

For making the adjustment we take that class which has lowest class-interval and adjust the frequencies of other classes in the following manner. If one class-interval is twice as wide as the one having lowest class-interval we divide the height of its rectangle by two, if it is three times more we divide the height of its rectangle by three, etc., i.e., the heights will be proportional to the ratio of the frequencies of the width of the class.

(i) When only mid points are given

When only mid points are given, ascertain the upper and lower limits of the various classes and then to construct the histogram in the same manner as discussed above.

Example: Draw the histogram for the following data:

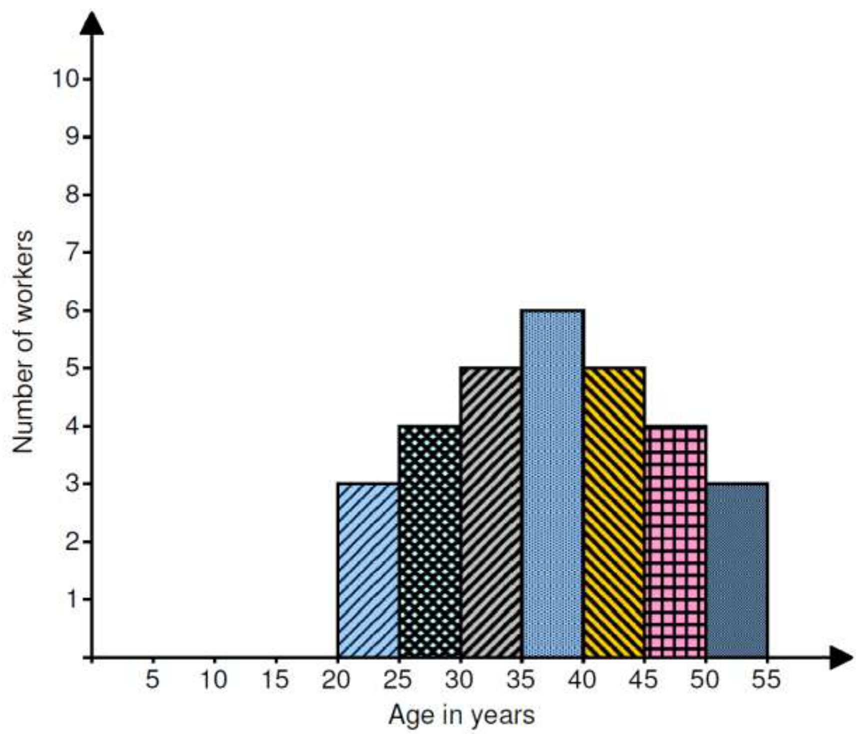
Age(in years)	20-25	25-30	30-35	35-40	40-45	45-50	50-55
Number of workers	3	4	5	6	5	4	3

Solution: A histogram (or rectangular diagram or block diagram) is a graphical representation of a frequency distribution in the form of rectangles one after the other with height proportional to the frequencies.

It is two dimensional in which case the height as well as width of the rectangle matters.

Since the first class is given to be 20-25 and the class interval is given to be 5, the remaining classes will be 25-30, 30-35 …. And soon. While preparing the frequency table it is to be kept in mind that a class, say35-40 and over but less than 40 and soon.

Since all the class interval are of equal magnitude, for drawing the histogram we construct rectangles on the class interval with height proportional to the corresponding frequencies of the classes.



3.6 FREQUENCY POLYGON

A frequency polygon can be constructed for a grouped frequency distribution, with equal-interval, in two different ways:

Method I:

Represent the class-marks along the x-axis.

Represent the frequencies along y-axis.

Join these points, in order, by straight lines.

The points at each end is joined to the immediate higher(or lower) class mark at zero frequency so as to complete the polygon.

Method II:

Represent a histogram of the given data.

Join the mid points of the tops of the adjacent rectangles by straight lines.

The mid points at each end are joined to the immediate higher (or lower) at zero frequency so as to complete the polygon.

The two classes, one at each end, are to be included

There are two ways of constructing a frequency polygon.

1. Draw a histogram of the given data and then join by straight lines the mid-points of the upper horizontal side of each rectangle with the adjacent ones. The figure so formed is called the frequency polygon. Polygon is closed by extending its both ends to the base line. When this is done two hypothetical classes at each end is included-each with a frequency of zero. This is done with the objective of making the area under polygon equal to the area under the corresponding histogram.
2. Another method is to take the mid points of the various class-intervals and then plot the frequency corresponding to each point and to join all these points by straight lines. This method is quite simple and time saving.

Advantages of frequency polygon over the histogram:

1. The frequency polygons of several distributions may be plotted on the same axis thereby making certain comparisons possible, whereas different histograms can't generally be plotted on the same axis. To compare histograms we must have a separate graph for each distribution.
2. The frequency polygon is simpler to construct than histogram.
3. It sketches an outline of the data pattern more clearly.
4. The polygon becomes increasingly smooth and curve-like as we increase the number of classes and number of observations.

   However in its construction same problems are faced as with histograms that is they cannot be used for distributions having open-end classes and suitable adjustment, as in the case of histogram, it is necessary when there are unequal class-intervals.

Example: Construct a frequency polygon for the following data:

Monthly pocket expenses of a student	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Number of students	10	16	30	42	50	30	16	12

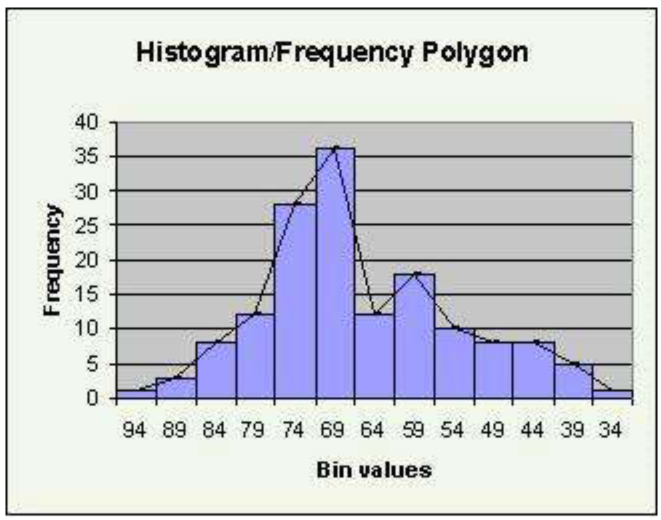
Solution:

Monthly pocket expenses of a student(in $)	Class-marks	Number of students
0-5	2.5	10
5-10	7.5	16
10-15	12.5	30
15-20	17.5	42
20-25	22.5	50
25-30	27.5	30
30-35	32.5	16
35-40	37.5	12

FREQUENCY POLYGON :(***Method II)***

The mid-points of the adjacent tops of rectangles of a histogram are joined by the line segments. The mid-point of each end is joined to the immediately lower or higher mid-points at x-axis. Thus we obtain a polygon which is called frequency polygon. Frequency polygons are useful for comparing distributions. This is achieved by overlaying the frequency polygons drawn for different data sets.

A frequency polygon is shown below:



Construction of smoothed frequency curve:

For drawing a smoothed frequency curve it is necessary to first draw the polygon and then smoothed it out. Construction of polygon has been discussed before. While polygon can be drawn even without first constructing histogram but smoothing of polygon can't be done properly without a histogram. Hence first construct a histogram then a polygon and then smooth the points of the polygon to obtain smoothed frequency curve.

As we know for constructing we will have to construct histogram first then the polygon then we will smooth the polygon curve to get the required smoothed curve. Figure shown below shows the required curve:

3.7 OGIVE:

Sometimes we are interested in knowing things like 'how many students have secured more than 50 marks' or 'how many workers earn more than Rs.1000 per month' etc. To answer these questions, it is necessary to add the frequencies. When frequencies are added, they are called cumulative frequencies. The curve obtained by plotting cumulative frequencies is called a cumulative frequency curve or an Ogive (pronounced Ojive).

Construction of ogive:

There are two methods of constructing ogive, namely

(a) The 'less than' method, and

(b) The 'more than method.

(a) Less than method:

In the 'less than' method we start with the upper limits of the classes and go on adding the frequencies. Thus we get less than cumulative frequencies. Then take the class intervals or the variate values in case of discrete distribution along the x-axis and plot the corresponding cumulative frequencies along the y-axis. The curve so obtained on joining the points by means of free hand drawing is called the less than cumulative frequency curve or less than ogive.

(b) More than method:

In this method we start with the lower limits of the classes and from the frequencies we subtract the frequency of each class. Then likewise the less than curve, class intervals or variate values are taken along x-axis and cumulative frequencies are plotted on y-axis, joining these points by free hand gives us more than ogive curve.

Cumulative frequency curve:

The Cumulative frequency curve for a grouped frequency distribution is obtained by plotting the points and then joining them by a free-hand smooth curve.

This is also known as ***ogive***.

Method:

Form the cumulative frequency table.

Mark the upper class limits along the x-axis.

Mark the cumulative frequencies along the y-axis.

Plot the points and join them by a free-hand smooth curve.

Example: Draw a cumulative frequency curve for the following data:

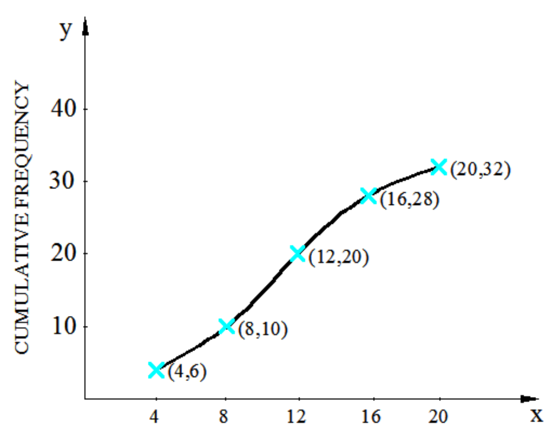
Marks	0-4	4-8	8-12	12-16	16-20
Number of students	4	6	10	8	4

Solution:

The cumulative frequency table is as follows:

Marks	*Number of students*	*Cumulative frequency*
0-4	4	4
4-8	6	4+6=10
8-12	10	10+10=20
12-16	8	20+8=28
16-20	4	28+4=32
Total	32	

Joining these points by a free-hand smooth curve, we have the following cumulative frequency curve:



We will take the marks-group (class intervals) along x axis and plot the less than frequencies along the y-axis against the corresponding upper limits of the corresponding classes. The smooth curve so formed is called 'less than' ogive.

Similarly if we plot more than cumulative frequencies against the lower limits of the corresponding classes we get 'more than' ogive.

To locate graphically the value of median, mark a point corresponding to ½ N along y-axis. At this point draw a line parallel to x-axis. From the point where it touches the ogive, draw a line perpendicular to x-axis meeting in 'M' say. Then abscissa of 'M' gives the value of median.

Utility of ogives:

From the view point of graphical representation, the ogive is especially used for the following purposes:

- To determine as well as to portray the number or proportion of cases above or below a given value.
- To compare two or more frequency distributions.
- Certain values such as median, quartiles, deciles etc. can be determined graphically using ogives.

SUMMARY:

In this lesson, we have discussed the following points:

Represent the frequency distribution of a character by means of a diagram: Histogram, frequency polygon, Frequency curve. Some basic definitions based on frequency distribution, rules for making a grouped frequency distribution.

3.8 SELF ASSESSMENT QUESTION:

1. Explain with suitable examples the following terms:

   a) Class limits   b) Frequency

   c) Cumulative frequency

2. Explain the various types of graphs for a frequency distribution .

3. Draw a cumulative frequency polygon for the following data:

Age(yrs)	Frequency
Less than 2	3
2-4	11
4-6	20
6-8	12
8-10	5

4. Represent the following distribution by a histogram :

Marks	No. of children
50-59	10
60-69	18
70-79	40
80-89	55
90-99	25
100-109	60
110-119	15

5. Given below are the marks obtained in mathematics by 30 students:

43 34 45 32 87 52 65 40 54 62 81 52 40 78 71 51 35 27 74 25 48
66 17 55 29 74 85 40 39 64

a) Construct a frequency distribution with class intervals of 10 marks.

b) Draw Ogives and hence find the median.

c) Find the proportion of students securing more than 75 marks

d) Find the marks below which 25% of the students have secured.

****

Unit - 1 **LESSON 4**

MEASURES OF LOCATION

STRUCTURE:

4.1 Objectives

4.2 Introduction

4.3 Arithmetic Mean

4.4 Geometric Mean

4.5 Harmonic Mean

4.6 Merits and demerits

4.7 Summary

4.8 Self assessment Exercises

4.1 OBJECTIVES:

The following are the main objectives of this lesson:

· To introduce the concept of measures of location.

· To know the arithmetic, geometric, and harmonic means

· To learn about the computational methods A.M.,G.M. and H.M.

· Derive and use some algebraic properties of these measures of central tendency.

4.2 INTRODUCTION:

One of the main characteristics of a frequency distribution is central tendency. The tendency of the observations to concentrate around a central point in a series is known as central tendency.

The measures which tell us the location or position of a central point in series are known as measures of locations or measures of position or measures of central tendency. The measures of central tendency are often called averages. Thus, averages are the values around which other items of the distribution congregate. The average represents a whole series and as such, its value always lies be the minimum and maximum values and is generally located in the centre or middle of the distribution.

The averages in this lesson are arithmetic mean, Geometric mean, and Harmonic mean. Let us consider one by one.

According to Prof.Yule a good average should posses the following characteristics.

It should be rigidly defined i.e., definition should be clear and unambiguous so that it leads to one and only one interpretation by different persons.

It should be easy to understand and calculate even for a non mathematical person.

It should be based upon all the observations.

It should be suitable for further mathematical treatment i.e., it should possess some important properties so that its use in statistical theory is enhanced.

It should affect least by fluctuations in the sampling.

It should not be affected much by the extreme values i.e., a few very small or very large observations should not unduly affect the value of a good average.

There are five measures of central tendency or measures of location which are commonly used in practice.

1. Arithmetic mean or Mean

2. Geometric Mean

3. Harmonic Mean

4.3 ARITHMETIC MEAN OR MEAN:

For a set of observations it is the sum of observations divided by the total number of observations. If $X_1, X_2 ....X_n$ are ***n*** observations, then heir mean is denoted by $\overline{X}$ is given by

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

In the case of frequency distribution having n observations $X_1, X_2 ....X_n$ with corresponding frequencies $f_1, f_{2,} f_3 ....f_n$. Arithmetic mean is given by

$$\overline{X} = \frac{f_1 X_1 + f_2 X_2 + ... + f_n X_n}{f_1 + f_2 + ... + f_n} = \frac{1}{\sum_i f_i}\sum_{i=1}^{n} f_i X_i = \frac{1}{N}\sum_{i=1}^{n} f_i X_i$$

Where $N = \sum_{i=1}^{n} f_i$ is the total of frequency

If the frequency distribution is grouped or continuous then 'X' is taken as the mid value of the corresponding class interval.

For values of 'X', calculations can be reduced by taking deviations from some arbitrary point 'A'. Then A.M is given by

$$\overline{X} = A + \frac{1}{N}\sum_{i=1}^{n} f_i d_i$$ **(Assumed mean method)**

Where $d_i = X_i - A$ and A is the value usually taken from middle part of the distribution.

In case pf continuous distribution arithmetic can be reduced by taking $d_i = \frac{X_i - A}{h}$, where h being the magnitude of the class interval. Then

$$\overline{X} = A + \frac{1}{N}\sum_{i=1}^{n} f_i d_i \times h$$ **(Step Deviation method)**

Merits

It is rigidly defined.

It is easy to understand and calculate.

It is based upon all the observations.

It is suitable for further mathematical treatment.

Of all averages A.M is affected least by fluctuations in sampling that is why it is called stable average.

Demerits

It cannot be located by inspection or graphically.

A.M cannot be used while dealing with qualitative characteristics.

It cannot be obtained if even a single observation is missing or lost unless we drop it out and calculate mean of the remaining values.

In case of extreme values it gives distorted picture.

Properties of arithmetic mean

1. Algebric sum of deviations of a given set of observations from their arithmetic mean is zero. i.e, if we are having n observations $X_1, X_2 ....X_n$ with corresponding frequencies $f_1, f_2, f_3 ....f_n$. Then

$$\sum_{i=1}^{n} f_i (X_i - \overline{X}) = 0 \text{ or for } X_1, X_2 ....X_n \text{ observations } \sum_{i=1}^{n} (X_i - \overline{X}) = 0$$

Proof: We have $\sum_{i=1}^{n} f_i (X_i - \overline{X}) = \sum_i f_i X_i - \sum_i f_i \overline{X}_i$ .... (1)

By definition $\frac{\sum_i f_i X_i}{\sum_i f_i} = \frac{\sum_i f_i X_i}{N} = \overline{X} \quad \Rightarrow N\overline{X} = \sum_i f_i X_i$ so from (1)

$$\sum_{i=1}^{n} f_i (X_i - \overline{X}) = N\overline{X} - N\overline{X} = 0 \quad \text{As } N = \sum_{i=1}^{n} f_i, \quad \text{hence proved.}$$

2. Sum of squares of deviations of a given set of observations is minimum when taken about mean

3. Effect of change in origin and scale of the arithmetic mean

Proof: Let X_i / f_i , $i = 1,2,3......n$ be a frequency distribution and suppose $d_i = X_i - A / h$ where *A* & *h* are arbitrary constants. Then mean of new distribution in the terms of original distribution have same transformation i.e., $\bar{d} = \bar{X} - A \big/ h$ **(called invariance property under linear transformations)**

Proof: We have $d_i = \frac{X_i - A}{h}$ or $\frac{\sum_i f_i d_i}{\sum_i f_i} = \frac{\sum_i f_i (X_i - A)}{h \sum_i f_i}$ Or

$$\bar{d} = \frac{\sum_i f_i X_i}{h \sum_i f_i} - \frac{A \sum_i f_i}{h \sum_i f_i} = \frac{\bar{X}}{h} - \frac{A}{h} = \frac{\bar{X} - A}{h} \quad \text{Or} \quad \bar{X} = A + h\bar{d}$$

Where *X-A* is change in origin & $\frac{X - A}{h}$ is the change in scale.

Hence A.M is independent of change in origin and scale

4. **Mean of a combined series :** If n_1 and n_2 are the sizes of means $\bar{X}_1$ & $\bar{X}_2$ are means of two groups. Then the mean $\bar{X}$ of the combined series is given by

$$\bar{X} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

Weighted mean

In some cases, some of the items are more important then the others this point should be borne in mined in order to make arithmetic mean true representative of the distribution. In such cases proper weightag should be assigned to the all the items in the distribution, the weights attached being proportional to the importance of the items

Let W_i be the weight attached to the X_i, *i=1, 2 ...n* item. Then weighted mean is given by

$$X_w = \frac{\sum_{i=1}^{n} X_i W_i}{\sum_{i=1}^{n} W_i}, \qquad ; \qquad i=1, 2 \ldots\ldots n.$$

Q: Find the weighted arithmetic mean of first n natural numbers

Answer: First n natural numbers are 1,2,3 ….n

Also 1+2+3+…..+n = $\frac{n(n+1)}{2}$ and $1^1+2^2+3^3+\ldots n^n = \frac{n(n+1)(2n+1)}{6}$ weighted arithmetic mean can be computed as follows

X: 1 2 3 ………n

W: 1 2 3 ………n

XW: 1^1 2^2 3^2 …… n^2

Now weighted mean

$$X_w = \frac{\sum_{i=1}^{n} X_i W_i}{\sum_{i=1}^{n} W_i} = \frac{1^1 + 2^2 + \ldots.. + n^2}{1 + 2 + \ldots. + n} = \frac{\frac{n(n+1)(2n+1)}{6}}{\frac{n(n+1)}{2}} = \frac{(2n+1)}{3}$$

Example 1: Find the arithmetic mean of the marks obtained by ten students of a class in mathematics in a certain examination. The marks obtained are :

25, 30, 21, 55, 47, 10, 15, 17, 45, 35

Solution: Let be the average marks.

Sum of all the observations =25+30+21+55+47+10+15+17+45+35=300

No. of students=10

∴ A.M. = 300/10=30

Example 2: Find the arithmetic mean from the following data ;

Marks	52	58	60	65	68	70	75
No. of students	7	5	4	6	3	3	2

Sol: Let x denotes the marks and *f* denotes No. of students in the frequency distribution of the table :

x	f	xf
52	7	364
58	5	290

60	4	240
65	6	390
68	3	204
70	3	210
75	2	150
Total	N = 30,	$\sum fx = 1848$

$$\text{A.M.} = \frac{\sum fx}{\sum f} = \frac{1848}{30} = 61.6$$

i) *Continuous series* :

If in a frequency distribution with n classes , be the mid points of the class intervals and f be the corresponding frequencies , then A.M. is given by

where $N = \sum f$

Example3: Calculate the Arithmetic mean for the following data:

C. I.	Freq.
10-20	2
20-30	7
30-40	17
40-50	29
50-60	29
60-70	10
70-80	3
80-90	2
90-100	1

Sol:

While calculating the arithmetic mean of such a tabular data , it is assumed that all the observations in any particular class interval have the same value .This value is the middle value or mid point of the class interval i.e. we replace the classes by the mid values.

C.I	Mid value (x)	Freq. (f)	fx
10-20	15	2	30
20-30	25	7	175
30-40	35	17	595
40-50	45	29	1305
50-60	55	29	1595
60-70	65	10	650
70-80	75	3	225
80-90	85	2	170
90-100	95	1	95
Total	N = 100	$\sum fx = 4850$	

Here N==100 , = 4850

$$\text{A.M.} = \frac{\sum fx}{\sum F} = \frac{4850}{100} = 48.50$$

Step deviation method:

When the class intervals in a grouped data are equal, then the calculate can be simplified further by taking out the common factor from the deviations. This common factor is equal to the width of the class-interval. In such cases, the deviations of variate x from assumed mean A are divided by the common factor .The Arithmetic mean is then obtained by the following formula

$$\text{A.M.} = A + \frac{\sum fd}{N}.i$$

Where A is the assumed mean

d = $\frac{\text{X-A}}{\text{è}}$, deviation of any variate from 'A'

i = Width of class interval

N = No. of observations.

Example 4: Calculate the average marks , by step deviation method , from the following data:

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No. of students	42	44	58	35	26	15

Sol: Let A=35 be the assumed mean.

Marks	No. of students(f)	Mid value(x)	d = $\frac{x-a}{i}$	fd
0-10	42	5	-3	-126
10-20	44	15	-2	-88
20-30	58	25	-1	-58
30-40	35	32	0	0
40-50	26	45	1	26
50-60	15	55	2	30
Total	N = 220		$\sum fd$ = -216	

Here N = = 220 and i = 10

Therefore, A.M., $= 35 + \frac{(-216)}{220} \times 10$ =25.18

4.4 GEOMETRIC MEAN:

Geometric mean is defined as the nth root of the product of n non-negative numbers is called geometric mean.

Formula:

i) *For individual series*

G.M.= $(x_1 \times x_2 \times x_3 \times .... \times x_n)^{\frac{1}{n}}$

ii) *For discrete distribution*

If the value (of the variable X) x(i=1,2,…….,n) with frequency f_i occurs (i=1,2,……,n) then,

$$\text{G.M.} = \frac{1}{N}\sum f_i(\log x_i)$$

Merits and demerits;

Merits:

1. It is rigidly defined
2. It is based on all the observations.
3. If G_1 *and* G_2 are the G.M. of two groups of sizes n_1 and n_2 respectively then the geometric mean G of the combined group of size $n_1 + n_2$ is given by

$$\text{LogG} = \frac{n_1 \log G_1 + n_2 \log G_2}{n_1 + n_2}$$

4. A.M. which has a bias for higher values, geometric mean has bias for smaller values and such is quit useful in phenomenon(such as prices)which has a lower limit (prices cannot go below zero) but has no such upper limit
5. It is not affected much by fluctuation of sampling.

Demerits:

1. It is not easy to understand to calculate for non mathematical person. If any one of the observations is zero, geometric mean becomes zero.

Example: Calculate geometric mean of the following data:

50 72 54 82 93

Sol:

X	Log X
50	1.6990
72	1.8573
54	1.7324

82	1.9138
93	1.9685
	$\sum \log X = 9.1710$

$$\text{LogG} = \frac{9.1710}{5} = 1.8342$$

G = antilog(1.8342) = 68.2653

Example: The following table gives the weight of 31 persons in a sample. Calculate G.M.

Weight(lbs) :	130	135	140	145	146	148	149	150	157
No. of persons :	3	4	6	6	3	5	2	1	1

Sol:

Weight, x	Frequency (f)	Log x	f log x
130	3	2.1139	6.3417
135	4	2.1303	8.5212
140	6	2.1461	12.8766
145	6	2.1614	12.9684
146	3	2.1614	6.4932
148	5	2.1703	10.8515
149	2	2.1732	4.3464
150	1	2.1761	2.1761
157	1	2.1959	2.1959
	N = 31	$\sum f \log x$	= 66.77

$$\text{Log G} = \frac{\sum f \log x}{N} = \frac{66.77}{31} = 2.1539$$

G.M. = antilog (2.1539)

= antilog (2.1539)

= 142.5 lbs

4.5 HARMONIC MEAN (H.M):

Harmonic mean, like geometric mean is a measure of central tendency in solving special types of problems.H.M. Is the reciprocal of the arithmetic mean of the reciprocal of the various observations of a variable Thus,

$$\text{H.M.} = (\text{for individual series}) = \frac{n}{\sum(1/x)}$$

$$\text{H.M. (in case of frequency)} = \frac{N}{\sum(f/x)}$$

Example : Calculate H.M. of the following data:

Marks	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Freq.	15	13	8	6	15	7	6

Sol: Computation of H.M.

Marks(x)	Mid point (m)	Freq. (f)	1/m	f/m
30-40	35	15	.02857	.4285
40-50	45	13	.02222	.28886
50-60	55	8	.01818	.14544
60-70	65	6	.01534	.09204
70-80	75	15	.01333	.19995
80-90	85	7	.01176	.08332
90-100	95	6	.01053	.6318
		N= 70		$\sum \frac{f}{m} = 1.86991$

$$\text{H.M.} = \frac{N}{\sum \frac{f}{m}} = \frac{70}{1.86991} = 37.4349$$

Merits and demerits of Harmonic mean:

Merits:

1) It is rigid and its value is always definite. No chance of computational bias.

2) It is also based on all the observations of the series.

3) It is capable of further algebraic treatment.

4) It is not affected by fluctuations of sampling.

5) It gives importance to the smaller values as such one big item cannot push up its value

6) It measures the relative changes and is extremely useful in averaging certain types of rates and ratios.

Demerits:

1) The computation of harmonic mean is not easy and is not readily understood.

2) It is not useful for the analysis of economic data as it gives weightage to items.

3) It is usually a value which doesn't exist in the series.

4) It is not a good representative of a statistical series, unless the phenomenon is such where small items have to be given a very weightage.

Properties of Harmonic Mean:

1) For two values a,b

A H = G^2

Proof: A = $\frac{a+b}{2}$, H = $\frac{2ab}{a+b}$

AH = $\frac{a+b}{2} \times \frac{2ab}{a+b}$ = G^2

2) Prove that the H.M. of a set of positive values is less than the G.M. of the same values than the A.M. i.e. or, the sign of equality holds iff all the observations are equal.

Proof: Let a and b be two real numbers a > 0 and b >0 then

A.M. = $\dfrac{a+b}{2}$ ……………. (1)

G.M. = $(ab)^{\frac{1}{2}}$ ……………. (2)

H.M. = $\dfrac{2ab}{a+b}$ ………….. (3)

Now, A.M. - G.M. = $\dfrac{a+b}{2}$ - $(ab)^{\frac{1}{2}}$ =1/2 $(\sqrt{a}-\sqrt{b})^2$, the sign of real quantity is always non negative, we have $(\sqrt{a}-\sqrt{b})^2 \geq 0$.

A.M-G.M ≥ 0 $\Rightarrow$ A.M $\geq$ G.M ………………… (4)

the sign of equality holds iff a = b

again,

$$\text{G.M-H.M}= \sqrt{ab} - \frac{2ab}{a+b} = \sqrt{ab}\left(\frac{a+b-2\sqrt{ab}}{a+b}\right) = \sqrt{ab}\,\frac{(\sqrt{a}-\sqrt{b})^2}{a+b} \geq 0$$

Since a >0, b >0 and square of real quantity is always non –negative.

Therefore G.M-H.M ≥ 0 ……………. (5)

The sign of equality hold iff. a=b

From (4) and (5) we have

A.M $\geq G.M \geq H.M$

The sign of equality holds iff. The two numbers are equal.

4.6 SELF ASSESSMENT:

1. Prove that the algebraic sum of deviation of a given set of observation from their a.m. is zero.

2. From the following data find the missing frequency when mean is 15.38

Size:	10	12	14	16	18	20
Frequency:	3	7	—	20	8	5

3. Define geometric mean and discuss its merits and demerits. Give two practical

situations where you will recommend its uses.

4. Define harmonic mean and discuss its merits and demerits. Under what situation would you recommend its use.

5. Calculate the A.M, G.M., H.M., of the following observations and A.M $\geq G.M \geq H.M$

   32,35,36,37,39,41,43,

****

UNIT - 1 **LESSON - 5**

MEDIAN AND MODE

STRUCTURE:

5.1 OBJECTIVES:

After studying this lesson, you should be able to:

* Compute median and mode from raw data or from a given frequency distribution.
* Derive and use some algebraic properties of measures of central tendency.

5.2 INTRODUCTION:

In probability and statistics a **median** is described as the numeric value separating the higher half of a sample, a population, or a probability distribution, from the lower half. The *median* of a finite list of numbers can be found by arranging all the observations from lowest value to highest value and picking the middle one. If there is an even number of observations, then there is no single middle value; the median is then usually defined to be the mean of the two middle values.

In a sample of data, or a finite population, there may be no member of the sample whose value is identical to the median (in the case of an even sample size), and, if there is such a member, there may be more than one so that the median may not uniquely identify a sample member. Nonetheless, the value of the median is uniquely determined with the usual definition. At most, half the population have values less than the *median*, and, at most, half have values greater than the median. If both groups contain less than half the population, then some of the population is exactly equal to the median. For example, if $a < b < c$, then the median of the list $\{a, b, c\}$ is b, and, if $a < b < c < d$, then the median of the list $\{a, b, c, d\}$ is the mean of b and c; i.e., it is $(b + c)/2$.

The median can be used as a measure of location when a distribution is skewed, when end-values are not known, or when one requires reduced importance to be attached to outliers, e.g., because they may be measurement errors. A disadvantage of the median is the difficulty of handling it theoretically.

5.3 MEDIAN:

Median is that value of the variable which divides a series into two equal parts so that one half or more of the items are equal to or less than it. This definition applicable in case of frequency distributions.

Easy explanation of the sample median

For an odd number of values

As an example, we will calculate the sample median for the following set of observations: 1, 5, 2, 8, 7.

Start by sorting the values: 1, 2, 5, 7, 8.

In this case, the median is 5 since it is the middle observation in the ordered list.

The median is the $((n + 1)/2)$th item, where n is the number of values. For example, for the list $\{1, 2, 5, 7, 8\}$, we have $n = 5$, so the median is the $((5 + 1)/2)$th item.

median = (6/2)th item

median = 3rd item

median = 5

For an even number of values

As an example, we will calculate the sample median for the following set of observations: 1, 5, 2, 8, 7, 2.

Start by sorting the values: 1, 2, 2, 5, 7, 8.

In this case, the average of the two middlemost terms is (2 + 5)/2 = 3.5. Therefore, the median is 3.5 since it is the average of the middle observations in the ordered list.

Formula:

i) *For individual series* : Let be the values of a variable written arranged in ascending order of magnitude; then the median is given by

a) If n is odd, then Median is the middle value after the observation have been arranged in ascending or descending order of magnitude.

b) If n is even, then Median =the a.m. of two middle terms observations after they are arranged in ascending or descending order of magnitude.

Example: Find the median of the following data:

X : 8, 10, 5, 9, 12, 11

Solution: In this data the number of items is 6 , which

Median = ascending order 5, 8, 9, 10, 11, 12, =Arithmetic mean of two middle terms

$$= \frac{1}{2}(9+10) = 9.5$$

ii) *For discrete series:*

Let be the n observations in ascending order with respective frequencies *f*. The median of this series/distribution is defined as the lowest value of the variable for which cumulative frequency (less than type) exceeds

Example: Calculate median for the following data:

No. of students	6	4	16	7	8	2
Marks	20	9	25	50	40	80

Solution: Arranging the marks in ascending order and preparing the following table:

Marks	frequency	Cumulative Freq.
9	4	4
20	6	10
25	16	26
40	8	34
50	7	41
80	2	43
Total	43	

Here N=43, Median = 22^{nd} value

The table shows that all items are from 10 to 25 have their values 25. Since 22^{nd} item lies in the interval, therefore, median is 25.

iii) For Grouped frequency distribution:

Here we, take median = Value, where N = .The lowest class for which the cumulative frequency (less than) exceeds is called median class. The value of the median is obtained under the assumption that in the Median classes are distributed uniformly throughout the interval by the following formula:

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

Where N = total frequency

i = magnitude of the median class

l = lower limit of the median class

f = frequency of the median class

C.f = cumulative frequency of the class preceding median class

Remarks: In case of grouped frequency distribution it is necessary to convert the class intervals into exclusive form if they are given in inclusive form.

Graphic method of locating median:

It can be easily located graphically with the help of a curve called the cumulative frequency curve or ogive.

Example: Find the median of the following data;

Class intervals:	0-6	6-12	12-18	18-24	24-30	30-36
Frequency :	5	11	25	30	23	18

Solution:

Class intervals	f	c.f<
0-5	5	5
5-10	11	16
10-15	25	41
15-20	30	71
20-25	25	94
25-30	18	112

Here N = 112

This value lies in the class interval 15-20, thus here f = 30, c.f.=41, i= 5 and hence

$$\text{Median} = 15+\frac{56-41}{30}\times 5 = 17.5$$

Hence median = 17.5

Hence we conclude that median value 17.5 lie in class interval 15-20

Merits and demerits of Median:

Merits :

1) It is rigidly defined.
2) It is easily computed and easily understood.
3) Unlike arithmetic mean, the median can be calculated where the data is incomplete. e.g., irregular class-intervals, or open class-intervals.

4) Its value is not affected by the items on the extreme.

5) It can be located merely by inspection in many cases

6) It can be located graphically.

7) It is best for qualitative data such as beauty, intelligence etc

Demerits:

1) Median may not be representative of a series in many cases. This is specially, when there are wide variations between the values of different items.

2) It is not suitable for further algebraic treatments.

3) Median cannot be calculated, if the frequencies of the class intervals are not uniformly spread over the values in the class interval.

4) Median is more likely to be affected by the fluctuations of sampling than Arithmetic mean.

5.4 MODE:

Mode is defined as that value in the series which occurs most frequently. In a frequency distribution, mode is that variate which has the maximum frequency. In other words, mode represents that value which is most frequent or typical or predominant.

Remarks:

1) There may be more than one mode in a frequency distribution .A frequency distribution with two modes is called bimodal, with three modes trimodal.

2) If each value occurs only once then there is no mode or all the values are modes.

Location of Mode:

1) For individual series:

a) If the number of observations is small we inspect mode at a glance by looking which one of the observation occurs most frequently.

b) If the number of observations is large we convert the individual series into discrete series and locate the mode by looking into the frequency table.

For example, in a series 6, 5, 3, 4, 7, 3, 8, 9, 5, 5, 4, we note that 5 occurs most frequently, therefore 5 is the mode.

2) *For discrete series* :

a) If there is a single mode we can locate mode at a glance by looking into the frequency column for the maximum frequency.

b) However, if the values cluster at more than one point or the frequencies are distributed irregularly then to find out the single value, we assume grouping method.

3) *For grouped series* :

a) If there is a single class with maximum frequency , we call this class as modal class and within this class mode is obtained by the formula:

$$\text{Mode} = L + \frac{F_M - F_{M-1}}{2F_M - F_{M-1} - F_{M-2}} \times i$$

Where

F_M = frequency of the modal class

F_{M-1} = frequency of the class preceding to modal class

F_{M-2} = frequency of the class following modal class

L = lower limit of modal class

i = width of class interval.

b) If the values cluster at more than one class interval we decide the modal class by grouping method and then use the formula given above. If this formula fails, then we use the following empirical formula

Mode = 3median-2mean

c) If all the class intervals are not of equal magnitude we convert the series into equal class intervals and then obtain the mode. If such a conversion is not possible then in case of moderately asymmetrical distribution, we use the formula

Mode = 3 Median - 2 mean.

d) We can locate mode graphically by drawing a Histogram.

Example: A Bata shop in Delhi had sold 100 pairs of shoes of Bata exclusive certain day with the following distribution:

Size of Shoe	4	5	6	7	8	9	10
No. of Pairs	10	15	20	35	16	3	1

Find the mode of the distribution.

Solution: let us prepare the table showing the frequency.

Size of Shoe	4	5	6	7	8	9	10
No. of Pairs	10	15	20	35	16	3	1

In the above table we notice that the size 7 has the maximum frequency, viz, 35. Therefore, 7, is the mode of the distribution.

Example: Find the mode of the following distribution.

Class	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	19	25	36	72	51	43	28

Solution: Computation of mode

Class	Frequency
0-10	19
10-20	25
20-30	36
30-40	72
40-50	51
50-60	43
60-70	28

By Looking into the frequency row we note that maximum frequency is 72.Hence the modal class is 30-40.Thus L = 30, $F_M = 72$, $F_{M-1} = 36$, $F_{M-2} = 51$, i=10.

$$\text{Mode} = 30 + \frac{72-36}{2 \times 72 - 36 - 51} \times 10 = 36.32$$

Hence we conclude that modal value 36.32 lie in the modal class 30-40

Example; Calculate the median and mode of the following:

Size:	0-10	10-20	20-30	30-40	40-50	50-60
Freq.:	4	16	15	20	7	5

Solution: Computation of median and mode

Size	frequency	C.F<
0-10	4	4
10-20	16	20
20-30	15	35
30-40	20	55
40-50	7	62
50-60	5	67

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

Here N/2 = 67/2 = 33.5, the c.f just> N/2 = 33.5. .Hence the corresponding class 20-30 is the median class. Using the median formula

$$\text{Md:} = 20 + \frac{33.5 - 20}{15} \times 10 = 29$$

Hence median is 2

Mode : here maximum frequency is 20. the corresponding class 30-40 is the model class. Using mode formula, we get

$$\text{Mode} = L + \frac{F_M - F_{M-1}}{2F_M - F_{M-1} - F_{M-2}} \times i$$

$$= 30 + \frac{20 - 15}{2 \times 20 - 15 - 7} \times 10 = 32.77$$

Hence mode is 32.77

Merits and demerits of Mode:

Merits:

1) It has the merit of simplicity. It can be obtained without much mathematical calculation. It can be located by inspection.

2) It is capable of being ascertained graphically.

3) It represents the most frequent value and hence it is very often in practice.

4) The arrangement of data is not necessary if the items are few.

5) Like median, mode is unaffected by the dispersion of the series.

Demerits:

1) It is not based on all the observations.

2) Mode is not rigidly defined, indeterminate, and indefinite.

3) Mode is not suitable for further mathematical treatment**.** For example, from the modal values and the sizes of two or more series, we cannot find the mode of the combined series.

4) As compared with mean, mode is affected to a greater extent by fluctuations of sampling.

Uses: Being the point of maximum density, mode is specially useful in finding the most popular size in studies relating to marketing, trade, business and industry. It is the appropriate average to be used to find the ideal size e.g., in business forecasting, in the manufacture of shoes or readymade garments, in sales, in production, etc.

5.5 SELF ASSESSMENT:

1) Following is the distribution of marks in law obtained by 50 students:

Marks more than	0	10	20	30	40	50
No. of students	49	46	40	20	10	4

Calculate the median mark.

2) Define median and give its merits and demerits and also to calculate median from the following given below data.

Wages (Rs.)	No. of laborers
60-70	3
70-80	5
80-90	10
90-100	20
100-110	8
110-120	2

3) The following table shows the age of distribution of persons in particular regions:

Age(yrs)	No. of persons
Below 10	18
10-20	24
20-30	32
30-40	42
40-50	55
50-60	35
60-70	25
70-80	10
Above 80	4

Calculate the appropriate measure of central tendency;

4) Define mode and also give its merits and demerits. To calculate the mode from the following frequency distribution:

Marks obtained	No. of candidates
0-9	6
10-19	29
20-29	87
30-39	181

40-49	247
50-59	263
60-69	133
70-79	43
80-89	9
90-99	2

5) Calculate mode by using empirical relation between mean, median and mode from the data given below:

Class interval	0-10	10-20	20-30	30-40	40-50	50-60
Frequency	4	20	35	55	62	67

****

UNIT - II **LESSON - 6**

MEASURES OF DISPERSION

STRUCTURE:

6.1 OBJECTIVES:

* The concept and purpose of dispersion
* Know the different coefficients of dispersion.
* Compute Range, inter quartile range and mean deviation from mean, median and mode of the data.
* Derive and use some algebraic properties of the dispersion.

6.2 INTRODUCTION:

It will be clear from the previous lesson that the measure of central tendency by itself can exhibit only one of the important characteristics of a distribution and, therefore while studying a distribution it is equally important to know how the variates are clustered around or scattered away from the point of central tendency .Thus if two groups of students have the same average marks, we may like to know whether one group consisting of students of average and near-average intelligence and the other group is made up of a large number of very bright and very dull students. Consider two distribution A and B:

A:	70	90	95	105	115	125
B:	15	20	25	70	190	280

Both the distributions have the same mean mean 100, yet they are different. If we are given that the mean of six observations is 100, we cannot form an idea as to whether it is the average of first series or second series. Thus we see that the measures of central tendency are inadequate to give us a complete idea of the distribution.

It will be clear from the above examples that although the arithmetic mean is same in the case of all the two distribution, yet the series widely differ from each other in their formation. The name given to this variation, deviation , scatter or spread is dispersion. Dispersion can, therefore, be defined as a "Lack of Uniformity" in the sizes or qualities of the items of a group or series. It may also be said that dispersion is the extent to which the magnitudes or the qualities differ, i.e., the degree of diversity. It must be noted that in measuring dispersion, the statistician is interested in the amount or the degree of the scatterdness, and not in its direction.

They must be supported and supplemented by some other measures. One such measure is Dispersion.

6.3 MEASURES OF DISPERSION:

Dispersion:

In a series of numerical values (or with in a group of the values of a variable) all the values are not uniform in their size (or magnitude) .Even the leaves of the same tree which look alike have great variation among themselves. Such a variation is inherent in nature. The word Dispersion is used in two senses in Statistics.

i) The scatter ness of the values of a variable due to variation among themselves is called Dispersion.

ii) The deviations from a measure of a central tendency or any other fixed value are not uniform in their size. The scatter ness of these deviations is also referred to as dispersion.

The terms dispersion give the idea of homogeneity or heterogeneity of the data under study.

Measures of Dispersion:

Since there are two meaning of dispersion into classes:

1) **Measures based on the Scatter ness of the values of a variable among themselves:**

   a) Range

   b) Interquartile Range

   c) Semi-Interquartile range or Quartile deviation

2) **Measures based on the spread of the deviations about some point**.

   a) Mean deviation

   b) Standard Deviation

The measures of dispersion are also called the "averages of second order". The requisites of a good measure of dispersion are same as given for the measures of central tendency (averages of first order). In the light of these essential requisites of satisfactory measure of dispersion we discuss the merits, demerits and uses of the various measures of dispersion.

Range:

The range is the length of the smallest interval which contains all the data. It is calculated by subtracting the smallest observation from the greatest and provides an indication of Statistical dispersion.

It is measured in the same units as the data. Since it only depends on two of the observations, it is a poor and weak measure of dispersion except when the sample size is large.

For a population, the range is greater than or equal to twice the Standard deviation, with equality only for the coin toss (Bernoulli distribution with p = ½).

The range, in the sense of the difference between the highest and lowest scores, is also called the crude range. When a new scale for measurement is developed, then a potential maximum or minimum will emanate from this scale. This is called the potential (crude) range. Of course this range should not be chosen too small, in order to avoid a ceiling effect. When the measurement is obtained, the resulting smallest or greatest observation will provide the observed (crude) range.

The midrange point, i.e. the point halfway between the two extremes, is an indicator of the central tendency of the data. Again it is not particularly robust for a small sample. The range is the simplest measure of dispersion .It is a rough measure of dispersion of dispersion. It is the difference between the maximum and minimum value. Thus

Range = largest value – Smallest value

Range = L – S

Range as calculated above is an absolute measure of dispersion which is unfit for the purposes of comparison if the distribution is in different units. For example, the range of the weight of students cannot be compared with the range of their heights measurements as the range of weights would be in Kgs. And the heights in centimeters. Thus sometimes for the purpose of comparison a relative measure of range is calculated:

Coefficient of Range = $\mathbf{X_{max} - X_{min}}$

Example : Find the range and the coefficient of range of the weights of 8 students from the following data;

27, 30, 35, 36, 38, 40, 40, 43.

Solution:

Range = L - S = 43- 27= 16

$$\text{Coefficient of range} = \frac{L-S}{L+S} = \frac{16}{70} = 0.23$$

6.4 MERITS AND DEMERITS OF RANGE

Merits:

1) It is the simplest measure of dispersion.

2) Its units are the same as the units of the variable being measured.

3) It is rigidly defined.

Demerits:

1) It is not reliable because it is affected by the extreme values.

2) It cannot be applied to open end classes.

3) Its sampling fluctuations are quite high in most cases and is rather difficult to ascertain.

4) It is not suitable for mathematical treatment.

Uses of Range:

1) Range is used in manufacturing units for the statistical quality control of the manufactured product by the construction of R-chart, i.e. the control chart of range.

2) Range is used in studying the variation in the prices of stock, shares, and other commodities that are sensitive to price changes from period to other period.

3) The meteorological department uses the range for weather forecasting since public is interested to know the limits within which the temperature is likely to vary on a particular day.

4) Range is of immense use in such kind of frequency distributions where the variations are not much.

2) Inter quartile range:

Definition: The difference between the third and the first quartile is called Inter quartile range.

Inter quartile range = Third quartile - first quartile.

Semi inter quartile range or Quartile Deviation:

Definition: Semi inter quartile range or Quartile Deviation is defined as the half the difference between the third and the first quartile .Symbolically,

Semi I.Q.R or $Q..D = \frac{Q_3 - Q_1}{2}$

It is better measure of dispersion as compared with **range**. It can be calculated for distributions with open ended classes

Since it utilizes 50% 0f the data but the other 50% remains unutilized, it cannot be regarded as reliable measures.

For comparative we use a relative measure of dispersion based on quartile deviation called as Coefficient of quartile devotion given by

$$Q..D = \frac{Q_3 - Q_1}{Q_3 - Q_1}$$

Example: Calculatle the quartile deviation and the coefficient of quartile deviation from the following data:

Age in years	10	20	30	40	50	60	70
No. of members	3	61	132	153	140	51	3

Solution:

Age in years	Freq.	c.f.
10	3	3
20	61	64
30	132	196
40	153	349
50	140	489
60	51	540
70	3	543

We know that: $Q..D = \frac{Q_3 - Q_1}{2}$

coefficient of $Q..D = \frac{Q_3 - Q_1}{Q_3 + Q_1}$

Q_3 = upper quartile

Q_1 = Lower Quartile

Where $Q_1 = \text{L} + \frac{N/4 - CF}{F} \times i$

$Q_3 = \text{L} + \frac{3N/4 - CF}{F} \times i$

Q.D. = 10

Coefficient of quartile deviation = 0.2

Merits, Demerits, and uses of quartile deviation:

Merits:

1). The quartile deviation is free from the objection attaching to range, i.e. it is not calculated from extreme items. Quartile deviation covers half of the items of the series. The 50 percentages items are those which are most likely to occur. The extreme 25 percentages items and the lower 25 percentages items are not considered and we take the middle 50 percentages items. If, for example, there are 1000 items in a series, quartile deviation will leave first 250 and the last 250 items and we consider the middle 500 items only.)

2) It is useful in studying dispersion in open end series, one in which the data may be ranked but measured qualitatively

3.) As compared to standard deviation, it is easy to calculate and simple to understand though not so simple as the range.

Demerits:

1) It is not capable of further mathematical treatment.

2) Its value is affected by sampling fluctuations.

3) It ignores 50 percentages of the items, i.e. it is not based on all the observations.

4) It doesn't take into account of the distribution below the first quartile or above the third quartile.

6.5 MEAN DEVIATION:

We have seen that the above methods of measuring dispersion suffer from common drawback namely, that they are not based on all the observation. To remove this difficulty another method of measuring dispersion is commonly employed and that is called mean deviation. This method is based on all the items of a series.

Definition: .

If X_i / f_i i=1,2,3…….n be a frequency distribution then the mean deviation about any arbitrary point A is given by

M.D(about any arbitrary point A)= $\frac{1}{N} \sum f|X - A|$

The arithmetic mean of the absolute deviations about any point 'A' is called the mean deviation or mean deviation about the point 'A'. Let be n variate-values, then are the deviations about any arbitrary point 'A'. These deviations may be positive or negative .If we ignore the negative sign of these deviations and consider it also as positive, then these are called absolute deviations. Mean deviation is not suitable for further mathematical calculation since it ignore negative sign.

Merits & Demerits of the Mean Deviation

Merits:

1. It is rigidly defined and based upon all the observations
2. It is easy to understand and calculate.
3. Since it is based upon deviation from some average, it provides better measure for comparison about the formation of different distributions.

Demerits:

Here the negative sign is being ignored which is mathematically unsound. It creates artificiality and renders mean deviation for further mathematical treatment.

It cannot be calculated for distributions with open ended classes.

Note: Mean Deviation is least when taken about median

6.6 COEFFICIENT OF MEAN DEVIATION:

To compare two or more groups of observations , we define coefficient of mean deviation .The quotient obtained by dividing the mean deviation by the point about which it is calculated is called coefficient of mean deviation .

$$\text{Coefficient of mean deviation} = \frac{Mean\,Deviation}{Median}$$

If mean deviation has been computed from mean or mode then Coefficient of mean deviation is equal to

$$\frac{Mean\,deviation}{mean} \quad \text{or} \quad \frac{Mean\,deviation}{\text{mod}\,e}$$

If frequency distribution mean deviation is computed by the formulae

$$\text{Mean Deviation } = \frac{\sum f|D|}{N}$$

$\sum f|D|$ = sum of the product of frequencies and deviation of items from the median ignoring $\pm$ signs.

Remarks:

1) Since sum of the deviations from mean is zero, we use the absolute mean deviation as a measured of dispersion.

2) When the value of arbitrary point A is not given, we calculate the mean deviation about arithmetic mean.

3) The mean deviation from median is preferred to mean deviation from mean because the median minimizes such deviation.

4) Sometimes we may calculate the median of the absolute deviations from the median and may call it the median deviation. In general mean deviation about median is equal to the quartile deviation if the distribution is symmetrical. The median deviation

would be preferred to the quartile deviation in a skewed series because it does accurately indicate the range around the median within which 50% of the observations fall.

Example: Calculate median and mean deviation about from the following given below data

Marks	:	0-10	10-20	20-30	30-40	40-50
No.of students	:	5	10	20	5	10

Solution: computation of median and mean deviation

Marks	f	c.f.	mid. Point	|D|=(X-25)	f|D|
0-10	5	5	5	20	100
10-20	10	15	15	10	100
20-30	20	35	25	0	0
30-40	5	40	35	10	50
40-50	10	50	45	20	200

Median = size of N/2th items, i.e., size of 50/2 = 25th items.

Median lies in class 20-30.

$$\text{Median} = L + \frac{N/2 - CF}{F} \times i$$

$$= 20 + \frac{25-15}{20} \times 10 = 25 \text{ marks}$$

$$\text{Mean deviation about median} = \frac{\sum f|D|}{N} = 9$$

Coefficient of mean deviation about median

$$= \frac{M.D, about\ median}{median} = \frac{9}{25} = .36$$

Example: Find out the mean deviation about mean in the following series.

Age(in yrs)	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of persons	10	25	22	40	32	35	10	8

Solution:

Class interval	Mid point, x	Freq	d = x-A	fd	$\lvert X-\overline{X}\rvert$	$f\lvert X-\overline{X}\rvert$
0-10	5	10	-30	-300	33.4	334
10-20	15	25	-20	-500	23.5	587.5
20-30	25	22	-10	-220	13.5	297
30-40	35	40	0	0	3.4	136
40-50	45	32	10	320	6.6	211.2
50-60	55	35	20	700	16.6	582
60-70	65	10	30	300	26.6	260.6
70-80	75	8	40	320	36.5	292
Total		N = 182		$\sum fd = 620$		$\sum f\lvert X-\overline{X}\rvert$ =2700.3

Mean, $= A + \frac{\sum fd}{N} = 35 + \frac{620}{182} = 38.40$

Mean deviation from mean $= \frac{\sum f\lvert X-\overline{X}\rvert}{N} = \frac{2700.3}{182} = 14.83$

Merits and Demerits of mean deviation:

Merits:

1) It is simple to understand and easy to compute.

2) It is not much affected by the fluctuations of sampling.

3) It is based on all the observations and gives weight according to their size.

5) It is rigidly defined.

6) It is flexible, because it can be calculated from any measure of central tendency.

Demerits:

1) It is not capable of algebraic treatment.

2) It gives more weight to the extreme deviations than the small ones.

3) It is very difficult and sometimes impossible to have any idea about its 'sampling fluctuations.'

4) Some artificiality is created due to the ignoring the signs of the deviations.

6.7 USES:

Mean deviation is rarely being used as a measure of variability. But due to ease and simplicity in the calculation it is rather useful in the business and Economics. It is very useful when the extreme deviations would influence standard deviation unduly. For example, to compute the personal distribution of wealth in a community, since extreme rich and poor are taken into consideration.

Example: Prove that mean deviation about mean of the variable x which takes the values $x_1, x_2 \ldots\ldots\ldots\ldots x_n$ *with frequency* $f_1 \; f_2 \ldots\ldots\ldots f_n$ is given by

$$\frac{2}{N}\left[\left[\bar{x}\sum_{x_i \angle \bar{x}} f - \sum_{x_i \prec \bar{x}} f_i x_i\right]\right]$$

Solution: Here x takes the values $x_1, x_2 \ldots\ldots\ldots\ldots x_n$ *with frequency* $f_1 \; f_2 \ldots\ldots\ldots f_n$ and $\sum f$ =N

Mean deviation about mean

$$\delta_{\bar{x}} = \frac{1}{N}\sum f_i |x - \bar{x}|$$

$$= \frac{1}{N}\left[\left[\sum_{x_i \angle \bar{x}} f_i(\bar{x} - x_i) + \sum_{x_i \geq \bar{x}} f_i(x_i - \bar{x})\right]\right] \ldots\ldots\ldots\ldots\ldots\ldots 1$$

We know that sum of the deviation about mean is zero i. e.

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

$$\text{or} \left[\left[\sum_{x_i \angle \bar{x}} f_i (x_i - \bar{x}) + \sum_{x_i \geq \bar{x}} f_i (x_i - \bar{x}) \right]\right] = 0$$

$$\sum_{x \geq \bar{x}} f_i \underset{i}{} (x - \bar{x}) = - \sum_{x \leq \bar{x}} f_i (x_i - \bar{x})$$

$$\therefore \quad = \sum_{x \leq \bar{x}} f_i (\bar{x} - x_i) \ldots\ldots\ldots\ldots\ldots\ldots\ldots 2$$

From (1) and (2), we have

$$= \frac{1}{N} \left[\left[\sum_{x_i \angle \bar{x}} f_i (\bar{x} - x_i) + \sum_{x_i \geq \bar{x}} f_i (x_i - \bar{x}) \right]\right]$$

$$= \frac{2}{N} \left[\left[\sum_{x_i \angle \bar{x}} f_i (\bar{x} - x_i) \right]\right]$$

$$= \frac{2}{N} \left[\left[\bar{x} \sum_{x_i \angle \bar{x}} f - \sum_{x_i \prec \bar{x}} f_i x_i \right]\right]$$

6.8 SELF ASSESSMENTS:

1) Explain the measure of dispersion? What purpose does a measure of dispersion serve?

2) Compute quartile deviation the following data:

Size	4-8	8-12	12-16	16-20	20-24	24-28	28-32	32-36
Freq.	9	13	18	25	15	12	10	7

3) Calculate mean deviation about mean from the following data:

x	10	20	30	40	50
f	11	14	26	13	10

4) Explain various method of measuring dispersion and point out their merits and merits.

5) Calculate the semi-interquartile range coefficient of quartile deviation:

Age in years	20	30	40	50	60	70	80
Freq.	3	61	132	153	140	51	3

6) Calculate the mean deviation about mean, median and mode from the following data:

Class intervals	Freq
0-9	2
10-19	11
20-29	15
30-39	42
40-49	35
50-59	24
60-69	13
70-79	3

****

UNIT - II **LESSONS - 7**

STANDARD DEVIATION AND COEFFICIENT OF VARIANCE

STRUCTURE:

7.1 OBJECTIVES:

After studying this lesson , you will be able to :

* Compute standard deviation of the data, whether grouped or ungrouped.
* Compute coefficient of all the measures of dispersion
* Derive and use some algebraic properties of standard deviation.
* Compute Quartile, Decile and percentiles from the data

7.2 INTRODUCTION:

In this lesson we shall discuss the concept of standard deviation, Standard deviation is a widely used measurement of variability or diversity used in statistics and probability theory. It shows how much variation or "dispersion" there is from the average (expected value). A low standard deviation indicates that the data points tend to be very close to the mean whereas high standard deviation indicates that the data are spread out over a large range of values.

Technically, the standard deviation of a statistical population data set, or probability distribution is the square root of its variance. It is algebraically simpler though practically less robust than the average absolute deviation A useful property of standard deviation is that, unlike variance it is expressed in the same units as the data.

In addition to expressing the variability of a population, standard deviation is commonly used to measure confidence in statistical conclusions. For example, the margin of error in polling data is determined by calculating the expected standard deviation in the results if the same poll were to be conducted multiple times. The reported margin of error is typically about twice the standard deviation – the radius of a 95 percent confidence interval In science researchers commonly report the standard deviation of experimental data, and only effects that fall far outside the range of standard deviation are considered statistically significant normal random error or variation in the measurements is in this way distinguished from causal variation. Standard deviation is also important in finance where the standard deviation on the rate of return on an investment is a measure of the volatility of the investment.

When only a sample of data from a population is available, the population standard deviation can be estimated by a modified quantity called the sample standard deviation

7.3 STANDARD DEVIATION AND VARIANCE:

The positive square root of the arithmetic mean of the squares of the deviations of the given observations in a series about its arithmetic mean is called standard deviation and is denoted by . σ

The standard deviation may be defined as the root mean square deviation when the deviations are taken about the arithmetic mean. In symbols:

i) Let the observations be $x_1\, x_2 ....,x_n$

then standard deviation,

$$\sigma = (\frac{1}{n}\sum(x_i - \bar{x})^2)^{\frac{1}{2}} \text{ ; where}$$

ii) Let the observations be $x_1\, x_2 ....,x_n$ with respective frequencies

$f_1 f_2 .... f_n$, then standard deviation,

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2} \quad ; \text{where and } \overline{x} = \frac{1}{N}\sum f_i x_i$$

iii) In case of grouped frequency distribution let $x_1 x_2 ...., x_n$ be the mid points of the class intervals with corresponding frequencies $f_1 f_2 .... f_k$, then

$$\text{Standard deviation, } \sigma = \left(\frac{1}{N}\sum f_i (x_i - \overline{x})^2\right)^{\frac{1}{2}}$$

$$\text{Where } \overline{x} = \frac{1}{N}\sum f_i x_i \text{ and N} = \sum f_i$$

Variance:

In probability theory and statistics, the variance is used as a measure of how far a set of numbers are spread out from each other. It is one of several descriptors of a probability distribution, describing how far the numbers lie from the mean (expected value). In particular, the variance is one of the moments of a distribution. In that context, it forms part of a systematic approach to distinguishing between probability distributions. While other such approaches have been developed, those based on moments are advantageous in terms of mathematical and computational simplicity.

The variance is a parameter describing in part either the actual probability distribution of an observed population of numbers, or the theoretical probability distribution of a not-fully-observed population of numbers. In the latter case a sample of data from such a distribution can be used to construct an estimate of its variance: in the simplest cases this estimate can be the sample

In general if X takes the values $x_1 x_2 ...., x_n$ with respective frequencies $f_1 f_2 .... f_n$ then variance is given as

$$\sigma^2 = \frac{1}{N}\sum f_i (x_i - \overline{x})^2 \quad \text{where N} = \sum f_i$$

7.4 STANDARD DEVIATION:

It is defined as the positive square root of the arithmetic mean of the squares of deviations of a given set of observations from their arithmetic mean. If $X_1, X_2 ....X_n$ be a given set n observation the standard deviation is denoted by σ and defined as

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

In the case of frequency distribution having n observations $X_1, X_2 ....X_n$ with corresponding frequencies $f_1, f_{2,} f_3 ....f_n$ standard deviation is given by

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i (X_i - \overline{X})^2} \quad , \qquad N = \sum_{i=1}^{n} f_i \quad \text{and} \qquad \overline{X} = \frac{1}{N}\sum_{i=1}^{n} f_i X_i$$

Merits & Demerits of the Standard Deviation

This measure satisfies almost all the criteria laid down for an ideal measure of dispersion. It is widely used measure of dispersion. Here the step of squaring $(X_i - \overline{X})$ removes the drawback of ignoring negative signs of deviations while computing mean deviation.

It gives greater weightage to extreme values as such hasn't found much favour with economist and businessmen who are more interested in the modal class.

Merits:

1) Standard deviation is rigidly defined.

2) Standard deviation is based on all the observations.

3) It is not much affected by sampling fluctuations.

4) It is not capable of further mathematical manipulations.

5) Standard deviation is the basis of various other Statistical measures. For example, correlation coefficient.

Demerits:

1) It is not easy to understand, and difficult to calculate .

2) It gives more weight to the extreme values, because the values are squared up.

3) It is affected by the value of every item in the series.

4) It is impossible to find it exactly in case of open end classes.

Uses:

Standard deviation is the best measure of dispersion. It is widely used in Statistics because it possess most of the characteristics of an ideal measure of dispersion. It is widely used in sampling theory and by biologists. It is used in the coefficient of correlation and in the study of symmetrical frequency distribution

7.5 PROPERTIES:

1. Standard deviation is independent of change of origin but not of scale.

Proof: Let us consider

$$d_i = X_i - A \quad \Rightarrow X_i = A + d_i \qquad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Multiply (1) both sides by f and summing over i=1,2,3,4.......n and then dividing by N we get

$$\frac{1}{N}\sum_{i=1}^{n} f_i X_i = \frac{A}{N}\sum_{i=1}^{n} f_i + \frac{1}{N}\sum_{i=1}^{n} f_i d_i \qquad \Rightarrow \bar{X} = A + \bar{d} \quad \ldots\ldots\ldots(2)$$

Subtracting (2) from (1)

$$(X_i - \bar{X}) = (d_i - \bar{d}) \qquad \ldots\ldots\ldots\ldots\ldots\ldots... (3)$$

Now $$\sigma_x{}^2 = \frac{1}{N}\sum_{i=1}^{n} f_i (X_i - \bar{X})^2 = \frac{1}{N}\sum_{i=1}^{n} f_i (d_i - \bar{d})^2 = \sigma_d{}^2$$

It means that variance and consequently standard deviation is independent of change in origin.

Again let $$d_i = \frac{X_i - A}{h} \quad \Rightarrow X_i = A + h d_i \qquad \text{where h >0}$$

$$\Rightarrow \bar{X} = A + \frac{h}{N}\sum_{i=1}^{n} f_i d_i = A + h\bar{d} \qquad \text{Or} \qquad (X_i - \bar{X}) = h(d_i - \bar{d})$$

Now $\sigma_x{}^2 = \frac{1}{N}\sum_{i=1}^{n} f_i (X_i - \overline{X})^2 = \frac{1}{N}\sum_{i=1}^{n} f_i \{h(d_i - \overline{d})\}^2 = h^2 \sigma_d{}^2$

Or $\sigma_x = h.\sigma_d$

It means that variance and consequently standard deviation is not independent of change in scale.

2. S.D is the minimum value of root mean square deviation.

It means s^2 will be least when deviations are taken from the mean i.e., when $\overline{X} = A$

Standard deviation of a combined series: If n1 and n2 are the sizes, $\overline{X}_1$ & $\overline{X}_2$ are means and σ_1 *and* σ_2 are standard deviations of two series respectively. Then combined standard deviation σ is given by

$$\sigma = \sqrt{\frac{1}{n_1 + n_2}\left[n_1(\sigma_1{}^2 + d_1{}^2) + n_2(\sigma_2{}^2 + d_2{}^2)\right]}$$

where $d_1 = \overline{X}_1 - \overline{X}$, $d_2 = \overline{X}_2 - \overline{X}$ and $\overline{X} = \frac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1 + n_2}$

How will you differentiate an absolute measure from a relative of dispersion?

Ans: If a measure of dispersion is expressed in the original units of the series, it is termed as absolutes measure of dispersion. A relative measure of dispersion is obtained by expressing as ratio or percentage, thus is a pure number independent of units or measurement. It is suitable for comparison of variability among two distributions.

Coefficient of dispersion:

In order to compare the variability of two series which differ widely in their averages or which are measured on different units of measurement. In addition to the measure of dispersion we calculate coefficient of dispersion. It is a pure number independent of units of measurement.

1. Coefficient based on range

C.D = $\frac{A - B}{A + B}$ where A and B are the largest and smallest values respectively.

2. Coefficient based on mean deviation

$$\text{C.D} = \frac{Mean\ \ Deviation}{average\ \ from\ \ which\ \ it\ \ is\ \ calculated}$$

3. Coefficient based on quartile deviation

$$\text{C.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

4. Based upon Standard deviation

$$\text{C.D} = \frac{\sigma}{\overline{X}}$$

Coefficient of Variation:

Defn. Coefficient of Variation

This measure was given by Prof. Karl Pearson Coefficient of variation:

100 times the coefficient of dispersion based upon standard deviation is called coefficient of variation (C.V.) i.e.

$$\text{C.V.} = \frac{\sigma}{\overline{X}} \times 100$$

In other words, C.V. is the percentage variation in the mean , standard deviation being considered as the total variation in the mean.

For comparing two series, we calculate the coefficient of variation for each series. The series having greater C.V. is said to be more reliable than the other and the series having lesser C.V. is said to be more consistent (or homogenous) than the other.

Uses:

1. It is widely used measure of dispersion
2. Smaller the value of C.V, more consistent is the data and vice-versa.
3. In the field experiments, It is generally reported low C.V indicates more reliability.

Example: Calculate the standard deviation and variance from the following data :

14, 22, 9, 15, 20, 17, 12, 11

Solution:

x	$(x-\overline{x})$	$(x-\overline{x})^2$
14	-1	1
22	7	49

9	-6	36
15	0	0
20	5	25
17	2	4
12	-3	9
11	-4	16
$\sum x = 120$		$\sum(x-\bar{x})^2 = 140$

$$\bar{x} = \frac{1}{n}\sum x = \frac{120}{15} = 15$$

$$\sigma^2 = \frac{1}{n}\sum(x_i-\bar{x})^2$$

$$= \frac{140}{15} = 17.5$$

Standard deviation, $\sigma = 4.18$

Variance = 17.5

Example: For any discrete distribution standard deviation is not less than mean deviation from mean, prove.

Solution:

Let the frequency distribution be as follows:

X: $x_1\, x_2 \ldots\ldots\ldots, x_n$ total

F: $f_1\, f_2 \ldots\ldots\ldots f_n$ N

We have to prove:

M.D not greater than S.D.

Or S.D. not less than M.D

Or S.D $\geq$ M.D

Or $(S.D.)^2 \geq (M.D)^2$

If $\frac{1}{N}\sum f_i(x_i-\bar{x})^2 \geq \left\{\frac{1}{N}\sum f_i|x_i-\bar{x}|\right\}^2$

Or if $\frac{1}{N}\sum f_i t^2{}_i \geq \frac{1}{N^2}(\sum f_i t_i)^2$ where $t_i = |x_i - \bar{x}|$

Or if $\frac{1}{N}\sum f_i t^2{}_i - (\frac{\sum f_i t_i}{N})^2 \geq 0$

Or if $\frac{1}{N}\sum f_i(t_i-\bar{t})^2 \geq 0$, $\bar{t} = \frac{1}{N}\sum f_i t_i$ which is true.

$\therefore$ S.D $\geq$ M.D

Example: Calculate standard deviation from the following data:

Marks :	10	20	30	40	50	60
No. of Students :	8	12	20	10	7	3

Solution: Let a = 30

Marks	f	d=x – 30	fd	fd ^{2}
10	8	-20	-160	3200
20	12	-10	-120	1200
30	20	0	0	0
40	10	10	100	1000
50	7	20	140	2800
60	3	30	90	2700
Total	N=60		$\sum fd = 50$	$\sum fd^2 = 10900$

Here N = 60,

Mean : $30 + \frac{50}{60} = 30.83$

Therefore, $\sigma^2 = \frac{1}{N}\sum fd^2 - (\frac{\sum fd}{N})^2$

$$= \frac{10900}{60} - (\frac{50}{60})^2$$

= 181.66-.6889 =180.97

$\sigma = 13.45$

Example: Compute standard deviation from the following data:

Class intervals	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Freq. (f)	8	12	17	14	9	7	4

Solution: Let a = 35, h = 10

Class Intervals	Mid point x	$d = \frac{x-a}{h}$	F	fd	fd^2
0-10	5	-3	8	-24	72
10-20	15	-2	12	-24	48
20-30	25	-1	17	-17	17
30-40	35	0	14	0	0
40-50	45	1	9	9	9
50-60	55	2	7	14	28
60-70	65	3	4	12	36
Total			N=71	$\sum fd$ =-30	$\sum fd^2$ =210

Here N= 71,

Mean $= A + \frac{\sum fd}{N} \times h = 35 + \frac{(-30)}{71} \times 10$ =30.77, and

$$\sigma^2 = \left[\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)\right]^2 \times h^2 =$$

$\Rightarrow \quad (2.96-.18)100 = 278$

σ = 16.67 Hence standard deviation is 16.67

Example : The run scored by two batsman A and B in various inning are :

A:	24	79	31	114	14	02	68	01	110	07
B:	05	18	42	53	09	47	52	17	81	56

Who is the better run getter? Who is more consistent?

Solution: Computation of coefficient of variation for batsman A:

X	x- $\overline{x}$	$(x-\overline{x})^2$
24	-21	441
79	34	1156
31	14	196
114	69	4761
14	-31	961
02	-43	1849
68	23	529
01	-44	1936
110	56	4225
07	-38	1444
Total=450		$\sum(x-\overline{x})^2$ =17498

$$\overline{x}_a = \frac{\sum x}{n} = \frac{450}{10} = 45$$

$$\sigma_A = \sqrt{\frac{17498}{10}}$$

= 41.83 runs

Coefficient if variation

$$V_A = \frac{\sigma_A}{\bar{x}_A}$$

$$= \frac{41.83}{45} = .9296 = 92.96\ \%$$

Or

$$V_A = \frac{\sigma_A}{\bar{x}_A} \times 100$$

$$= 92.96\%$$

Computation of coefficient of variation for batsman B:

X	$(x-\bar{x})$	$(x-\bar{x})^2$
05	-33	1089
18	-20	400
42	4	16
53	15	225
09	-29	841
47	9	81
52	14	196
17	-21	441
81	43	1849
56	18	324
380	_	5462

$$\bar{x}_B = \frac{\sum x}{n} = \frac{80}{10} = 38\text{Runs}$$

$$\sigma_B = \sqrt{\frac{5462}{10}} = 23.37$$

$\therefore$ Coefficient of variation

$$V_B = \frac{\sigma_B}{\bar{x}_B} = \frac{23.37}{38}$$

$= .615 = 61.50\%$

Conclusion: (1) Here $\bar{x}_A > \bar{x}_B$; hence batsman A is better run getter than B

(2) Here $V_A > V_B \; or \; V_B < V_A$

Thus the variability in terms in the runs scored by B is less than the variability in the runs scored by A. Hence B is more consistent player.

7.6 PARTITION VALUES

These are the values which divide the series into a number of equal parts.

Quartiles

The three points which divide the series into four equal parts are called quartiles. The first quartile, Q1, is the values which exceed 25% of the observations and is exceeded by 75% of the observations. The second quartile, Q2, coincides with median. The third quartile, Q3, is the point which has 75% observations before it and 25% observations after it.

The values of the variable x corresponding to $\left[\frac{N+1}{4}\right]^{th} \left[\frac{N+1}{2}\right]^{th} \left[\frac{3(N+1)}{4}\right]^{th}$ items of an ordered discrete series are the values of Q1, Q2 and Q3 respectively and their position easily be adjudged with the help of cumulative frequencies.

Formula for calculating ith quartile (i = 1, 2, 3) for a continuous distribution

$$Q\,i = l + \left(\frac{\frac{iN}{4} - C}{f}\right) \times h$$

Where

Q i = ith quartile which is to be calculated

h = magnitude or width of the quartile class

l = lower limit of the quartile class

f = frequency of the quartile class

C = cumulative frequency of the class just above the quartile class

N = total of frequency

Deciles

The nine points which divide the series into ten equal parts are called deciles Hence, there are in all nine deciles denoted D1, D2, .., D9.so, the number of items that lie between any two deciles is also 10 per cent. D5, the fifth decile divides the series into halves and hence, it is the same as median.

The values of the variable x corresponding to $\frac{i(N+1)}{10}^{th}$ items of an ordered discrete series wiil give the value of Di for i= 1,2…..9 respectively and their position easily be adjudged with the help of cumulative frequencies.

Formula for calculating ith decile (i = 1, 2, …9) for a continuous distribution

$$\text{Di} = l + \left(\frac{\frac{iN}{10} - C}{f} \right) \times h$$

Where

D i = ith decile which is to be calculated

h = magnitude or width of the decile class

l = lower limit of the decile class

f = frequency of the decile class

C = cumulative frequency of the class just above the decile class

N = total of frequency

Percentiles

Percentiles are the ninety-nine points which divide the series into hundred equal parts. For example the forty-sixth percentile, is the point which exceeds 46% of the observations. The values of the variable x corresponding to $\frac{i(N+1)}{100}^{th}$ items of an ordered discrete series will give the value of Pi for i= 1,2…..99 respectively and their position easily be adjudged with the help of cumulative frequencies.

Formula for calculating ith percentile (i = 1, 2, …99) for a continuous distribution

$$Pi = l + \left(\frac{\frac{iN}{100} - C}{f} \right) \times h$$

Where Pi = ith percentile which is to be calculated

h = magnitude or width of the percentile class

l = lower limit of the percentile class

f = frequency of the percentile class

C = cumulative frequency of the class just above the percentile class

N = total of frequency

7.7 SELF ASSESSMENTS:

1). What is variability? Define and discuss the common measure of variability. Explain the circumstances, if any, in which you would advise the ues of each.

2). The following table gives the distribution of wages in the two branches of a factory:

Monthly Wages (Rs.	No. of workers	
	BranchA	Branch B
100-150	167	36
200-250	205	93
250-300	253	157

300-350	205	105
	168	82
Total	1000	500

Find the mean and standard deviation for the two branches for the wages separately.

a) Which branch pays higher average wages?

b) Which branch has greater variability in the wages in the relation to the average wages?

c) What is the average monthly wage for the factory as a whole?

d) What is the variance of wages of all of the workers in the two branches A and B taken together?

4) What is standard deviation? Explain the superiority of standard deviation on the other measure of variability.

5) The following is the record of number of bricks laid each day 10 days by two bricks - layers A and B. Calculate the coefficient of variation in each case and discuss the relative consistency of the two bricks- layers:

A	700	675	725	625	650	700	650	700	600	650
B	550	600	575	550	650	600	550	525	625	600

****

UNIT - II **LESSON - 8**

MOMENTS

STRUCTURE:

8.1 OBJECTIVES:

After reading this lesson you will be able to :

- Understand the meaning of moments.
- Understand the moments about any arbitrary point and central point.
- Find out the different moments from the data.
- The limitations of moments.
- Know the concept and purpose of Sheppard's correction.

8.2 INTRODUCTION:

The various measures of location or central tendency have been discussed in previous lessons; there in you have also learnt how to compute these measures and their utilities. The present lesson deals about moments. Moment is a term generally used in Physics and refers to the turning effect or rotating effect of a force. When it is applied in Statistics, it describes the various characteristics of a frequency distribution, viz; central tendency, dispersion, skewness and kurtosis. With the help of moments we study the distribution.

8.3 MOMENTS:

Given n observations $x_1\ x_2.........,x_n$, and an arbitrary constant A,

$\frac{1}{n}\sum(x-A)$ is called the first moment about A,

$\frac{1}{n}\sum(x-A)^2$ is called the 2nd moment about A

$\frac{1}{n}\sum(x-A)^3$ is called the 3rd moment about A,

And soon. Let us denotes these moments successively by $m_1', m_2', m_3', etc.$

Then $\mu_1' = \sum\frac{(x-A)}{n} = \bar{x} - A$

i.e the first moment about A = $\bar{x} - A$

moment about zero and moment about mean are particularly important.

1) **Moment about zero (or Raw moment)**

Ist moment about zero $= \frac{1}{n}\sum x = \bar{x}$

2nd moment about zero $= \frac{1}{n}\sum x^2$

3rd moment about zero $= \frac{1}{n}\sum x^3$

And so on. Note that the first moment about zero is the mean $\bar{x}$.

$m' = \bar{x}$

2) Moment about mean (or central moments)

Ist moment about mean $= \frac{1}{n}\sum(x-\bar{x}) = 0$

2nd moment about mean $= \frac{1}{n}\sum(x-\bar{x})^2 = \sigma^2$

3rd moment about mean $= \frac{1}{n}\sum(x-\bar{x})^3$

4th moment about mean $= \frac{1}{n}\sum(x-\bar{x})^4$

And soon. These are usually denoted by $m_1, m_2\ m_3\quad m_4$ etc..(sometimes, these are represented by $\mu_1, \mu_2, \mu_3, \mu_4$ etc.). Note that the first central moment is always zero, and the 2nd moment is the variance.

m_1=0, $\qquad m_2 = \sigma^2$

From the second relation, we find that the standard deviation is the square root of the second central moment m_2.

The 3rd central moment m_3 is used to measure skewness and the 4th ventral moment m_4 is used to measure kurtosis. Higher order moments m_5, m_6 etc. are seldom used.

In general, given n observations $x_1, x_2, \ldots\ldots\ldots. x_n$, the r-th order, moments (r= 0,1, 2….) are defined as follow:

r-th moment about A: $m_r' = \frac{1}{n}\sum(x-A)^r$

if we put r = 1 ,2 and soon

$$m_1' = \frac{1}{n}\sum(x-A)^1$$

$m_2' = \frac{1}{n}\sum(x-A)^2$ and soon.

r-th central moment: $m_r = \frac{1}{n}\sum(x-\bar{x})^r$

for a frequency distribution,

r-th moment about A: $m'_r = \frac{1}{n}\sum f(x-A)^r$ (raw moment)

r-th central moment: $m_r = \frac{1}{n}\sum f(x-\bar{x})^r$

where N= $\sum f$.(Note that moment about mean are written without dashes, but moments about any other origin, i.e. non central moments, with dashes)

there are important relation between central and non central moment. For example, if the non-central moments($m'_1, m'_2, m'_3, etc.$) about any arbitrary origin A are known, the central moments can be obtained by using the relation

$$m_2 = m'_2 - m'^2_1$$

$$m_3 = m'_3 - 3m'_2 m'_1 + 2m'^3_1$$

$$m_4 = m' - 4m'_3 m'_1 + 6' m_2 m'^2_1 - 3m'^4_1$$

In particular, using the first two moments m'_1 and m'_2 , about an arbitrary origin A the mean and variance may be obtained:

$$\bar{x} = m'_1 + \text{A}, \quad m_2 = m'_2 - m'^2_1 = \sigma^2$$

Remarks:

1. The first moment of order one about origin is equal to mean
   i.e. $\mu'_1 = \bar{x}$
2. The first moment about mean is always zero.

$$\mu_1 = \frac{1}{N}\sum f(x-\bar{x}) = 0$$

3. The second moment about mean which is equal to variance.

$$\mu_2 = \frac{1}{N}\sum f(x-\overline{x})^2$$

4. If the distribution is symmetrical about arithmetic mean, then all the odd order moments are zero.

5. The moments of order zero are always to 1.

Limitations of Moments:

1. Since there are four main characteristics of the frequency distributions which can be measured by the first four moments completely, therefore we do not study the moments of higher order.

	Moment:	characteristics of the distribution
i)	First moment about origin	Central tendency
ii)	Second moment about mean	Variability
iii)	Third moment about mean	Skewness
iv)	4th moment about origin	Kurtosis

2. Moments of higher order, though important in theory are extremely sensitive to sampling fluctuations

3. In case of higher moments, Charlier's check becomes difficult.

4. The assumption of normality is rarely found in economics and social studies in practice. So comparison by moments is not appropriate.

5. Moments do not exist for some theoretical distributions.

6. In case of symmetrical distribution all odd order moments are zero.

Effect of Change of Origin and Scale oncentral Moments:

Let x: $x_1, x_2, \ldots\ldots\ldots x_n$

f: $f_1, f_2 \ldots\ldots\ldots f_n$

be a frequency distribution.

Let $\mu = \dfrac{\overline{x}-a}{h}$ or $\overline{x} = a + h\overline{u}$.

Let μ_r = r-th moment about mean for x $= \frac{\sum f(x-\bar{x})^r}{\sum f}$

And m_r =r-th moment about mean $\bar{\mu}$ for $\mu = \frac{\sum f(\mu-\bar{\mu})^r}{\sum f}$

Now $\quad m_r = \frac{\sum f(\mu-\bar{\mu})^r}{\sum f}$

$$= \frac{1}{N}\sum f\left(\frac{x-a}{h} - \frac{\bar{x}-a}{h}\right)^r$$

$$= \frac{1}{\sum f}\sum f\left(\frac{x-\bar{x}}{h}\right)^r = \frac{1}{h^r}\mu_r \text{, r=1 , 2, 3, 4…..}$$

Or $\quad \mu_r = h^r m_r$

$\Rightarrow$ r th central moment of the variable x is equal to h^r times , r th central moment of the variable μ_r where $\mu = \frac{x-a}{h}$.

Thus we note that central moments are independent of the change of origin but not of scale. i.e. the moments about mean are in variant for the change of origin but not of scale.

Effect of change of origin and scale on non-central moments

i) $\quad \mu_r'(a)$ = r th moment about a for x = $a + hui$

$$= \frac{1}{N}\sum f_i(a+hu_i-a)^r \text{ where } u_i = \frac{x_i-a}{h}$$

$$= \frac{h^r}{N}\sum f_i u_i^r$$

$= h^r$ (r th moment of u about origin)

$\Rightarrow$ r th moment about a for the variable x is h^r times the r th moment about origin (zero) of the variable u.

ii). μ'_r = r th moment about zero for the variable x

$$= \frac{1}{N}\sum f_i x_i^{\,r} \quad N = \sum f$$

$$= \frac{1}{N}\sum f_i (a + hu_i)^r \quad \text{where } u_i = \frac{x_i - a}{h}$$

$$= \frac{1}{N}\sum f_i \left(a^r + rC_1 a^{r-1}(hu_i) + rC_2 a^{r-2}(hu_i)^2 + ... + (hu_i)\right)$$

$$= a^r + rC_1 a^{r-1} h m'_1 + rC_2 a^{r-2} h^2 m'_2 + ... + h^r m'_r .$$

$\Rightarrow$ r th moment of x about origin can be expressed in terms of the r th and lower moments of u about origin. Hence μ'_r is not independent of the change of origin and scale.

Moments about a in terms of moments about any other point b

Let $\mu'_r(a)$ = r th moment about a

$\mu'_r(b)$ = r th moment about b

And for frequency distribution x: $x_1, x_2, ..., x_n$ total

f: $f_1, f_2 ...., f_n$ $N=\sum f$

Now $\mu'_r(a) = \frac{1}{N}\sum f_i (x_i - a)^r$

$$= \frac{1}{N}\sum f_i (x_i - b + b - a)^r \text{ , adding and subtracting b}$$

$$= \frac{1}{N}\sum f_i \left[(x_i - b)^r + rC_1 (x - b)^{r-1}(b - a) + ...rC_r (b - a)^r\right]$$

$$= \frac{1}{N}\sum f_i (x_i - b)^r + rC_1 (b - a)\frac{1}{N}\sum f_i (x_i - b)^{r-1} + ... + (b - a)^r$$

$$= \mu'_r(b) + rC_1(b-a)\mu'_{r-1}(b) + ... + (b-a)^r$$

In particular

$$\mu'_1(a) = \mu'_1(b) + (b-a)$$

$$\mu'_2(a) = u'_2(b) + 2(b-a)u'_1(b) + (b-a)^2$$

$$\mu'_3(a) = = u'_3(b) + 3(b-a)u'_2(b) + 3(b-a)^2 u'_1(b) + (b-a)^3$$

$$\mu'_4(a) = = u'_4(b) + 4(b-a)u'_3(b) + 6(b-a)^2 u'_2(b) + 4(b-a)^3 u'_3(b) + (b-a)^4$$

Remark. For simplicity we may write b-a = d

8.4 RELATION BETWEEN MOMENTS ABOUT MEAN IN TERMS OF MOMENTS ABOUT ANY POINT:

We have

$$\mu_r = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^r \qquad = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - A + A - \bar{x})^r = \frac{1}{N}\sum_{i=1}^{n} f_i(d_i + A - \bar{x})^r = \frac{1}{N}\sum_{i=1}^{n} f_i(d_i - \mu'_1)^r$$

[Since for $(x_i - A) = d_i$ we have $\bar{x} = A + \frac{1}{N}\sum_{i=1}^{n} f_i d_i{}^r = A + \mu'_1$ as $\mu_r{}' = \frac{1}{N}\sum_{i=1}^{n} f_i d_i{}^r$]

$$= \frac{1}{N}\sum_i f_i\,(d_i^r - {}^rC_1\, d_i^{r-1}\, \mu_1' + {}^rC_2\, d_i^{\,r-2}\, \mu_1'^2 - {}^rC_3\, d_i^{\,r-3}\, \mu_1'^3 + \dots + (-1)^r\, \mu_1'^r\,].$$

$$= \mu_r' - {}^rC_1\mu_{r-1}'\, \mu_1' + {}^rC_2\, \mu_{r-2}'\, \mu_1'^2 - \dots + (-1)^r\, \mu_1'^r$$

In particular, on putting r = 2, 3 and 4 in above equation and simplifying, we get

$$\left.\begin{aligned} \mu_2 &= \mu_2' - \mu_1'^2 \\ \mu_3 &= \mu_3' - 3\mu_2'\, \mu_1' + 2\mu_1'^3 \\ \mu_4 &= \mu_4' - 4\mu_3'\, \mu_1' + 6\mu_2'\, \mu_1'^2 - 3\mu_1'^4 \end{aligned}\right\}$$

8.5 RELATION BETWEEN MOMENTS ABOUT ANY POINT IN TERMS OF MOMENTS ABOUT MEAN: WE HAVE

$$\mu_r{}' = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - A)^r = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x} + \bar{x} - A)^r = \frac{1}{N}\sum_{i=1}^{n} f_i(z_i + \mu'_1)^r$$

Where $(x_i - \bar{x}) = Z_i$ and $\bar{x} = A + \mu'_1$

Thus $\mu_r' = \frac{1}{N}\sum_i f_i (z_i^r + {}^rC_1 z_i^{r-1}\mu_1' + {}^rC_2 z_i^{r-2}\mu_1'^2 + \ldots + \mu_1'^r)$

$$= \mu_r + {}^rC_1\mu_{r-1}\mu_1' + {}^rC_2\mu_{r-2}\mu_1'^2 + \ldots + \mu_1'^r.$$

In particular, putting r = 2, 3, 4 and noting that $\mu_1 = 0$ we get

$$\mu_2' = \mu_2 + \mu_1'^2 \quad ,$$

$$\mu_3' = \mu_3 + 3\mu_2\mu_1' + \mu_1'^3$$

$$\mu_4' = \mu_4 + 4\mu_3\mu_1' + 6\mu_2\mu_1'^2 + \mu_1'^4$$

Example. The first four moments of a distribution about 5 are 2, 20, 40, and 150. Find the first four moments about origin (i.e. zero) and moment about mean.

Solution. Here we are given

$$u_1'(5) = 2,\ u_2'(5) = 20, \quad u_3'(5) = 40 \quad u_4'(5) = 150$$

$$\therefore\ u_1' = A + \frac{1}{N}\sum fx \quad = 5+2 = 7$$

$$u_2' = \frac{1}{N}\sum fx^2 = \frac{1}{N}\sum f((x-5)+5)^2$$

$$u_2' = 20+20+25=65$$

$$u_3' = \frac{1}{N}\sum fx^3 = \frac{1}{N}\sum f((x-5)+5)^3$$

$$u_3' = \quad 40+300+150+125=615$$

$$u_4' = \frac{1}{N}\sum fx^4 = \frac{1}{N}\sum f((x-5)+5)^4$$

$$u_3 = u_3' - 3u_2'u' + 2u_1'^3 \quad = 150+800+3000+1000+625=5575$$

$$u_4' = 5575$$

Moments about mean

u_1 = First moment about mean is always equal to zero.

$u_2 = u_2' - u_1'^2$ = 65-49=16

$u_3 = u_3' - 3u_2'u' + 2u_1'^3$ =615-3(65)(7)+2 (7^3)

615-1365+2(343)=615-1365+686

u_3 = -64

$u_4 = u_4' - 4u_3'u_1' + 6u_2'(u_1')^2 - 3(u_1')^4$

=5575-4(615)(7)+6(65)(49)-3(7^4)

=5575-17220+19110-7203

u_4 = 262

Example. The first four moments of a distribution about the value 0 are -209, 176, -236 and 1088. Find the moments about mean.

Solution. Given u_1'=-20, u_2' =176

u_3' = -236, u_4' = 1088

Therefore $u_2 = u_2' - u_1'^2$ =176-$(-20)^2$ =176-400<0

Since u_3 comes out to be negative, hence we conclude that this data is inconsistent and there is no use of calculating the moment of order 3 and 4.

Remarks:

In our opinion the data might have taken arbitrarily or by multiplying the values $m_1' = .20$, m_2' =1.76, m_3'=-2.36 , m_4' = 10.88 by 100. Such a multiplication is not justified because $u_r' = h^r m_r'$, r =1 , 2, 3, 4.

Example. Calculate the first four moments about mean for the following data.

X:	1	2	3	4	5	6	7	8	9
F:	1	6	13	25	30	22	9	5	2

Solution:

Calculation for moments

X	f	d= x-5	fd	fd^2	fd^3	fd^4
1	1	-4	-4	16	-64	256
2	6	-3	-18	54	-162	486

3	13	-2	-26	52	-104	208
4	25	-1	-25	25	-25	25
5	30	0	0	0	0	0
6	22	1	22	22	22	22
7	9	2	18	36	72	144
8	5	3	15	45	135	405
9	2	4	8	32	128	512
	113	0	-10	282	2	2058

Moments about the point x=5

d = x-a = x-5 (A=5) N =113

$$u_1' = \frac{\sum fd}{N} = \frac{-10}{113} = -.0885$$

$$u_2' = \frac{\sum fd^2}{N} = \frac{282}{113} = 2,4956$$

$$u_3' = \frac{\sum fd^3}{N} = \frac{2}{113} = .0177$$

$$u_4' = \frac{\sum fd^4}{N} = \frac{2058}{113} = 18.2124$$

Moments about mean

u_1 = First moment about mean is always equal to zero

$$u_2 = u_2' - u_1'^2 = 2.4878$$

$$u_3 = u_3' - 3u_2'u' + 2u_1'^3 = .0177\text{-}3(2.,4956)(\text{-}.0885)\text{+}2\,(-.0885)^3$$

$$u_3 = .6789$$

$$u_4 = u_4' - 4u_3'u_1' + 6u_2'(u_1')^2 - 3(u_1')^4$$

After simplification we get

u_4 =18.3357

Example: the first three moments of a distribution about the value 67 of the variable are .45, 8.73 and 8.91. Calculate the second and third central moments of the distribution.

(Do it yourself).

8.6 SHEPPARD'S CORRECTION:

In the computation of mean, standard deviation and moments etc., from a grouped frequency distribution, the fundamental assumption is that the mid-points are reasonable approximations to the means of observations within class intervals, i.e., the class frequencies concentrate at the mid-value. However, in case of symmetrical or moderately asymmetrical distributions, the true mean of an interval is actually closer to the general mean than the mid point. Thus the deviations from the mid point are too large. W.F. Sheppard suggested a correction for this error in the moments. These corrections are called Sheppard's correction for grouping. Sheppard's corrections for central moments are

i) $u_2 \text{ (Correct)} = u_2 \text{ (calculated)} - \frac{h^2}{12}$

ii) $u_3 \text{ (Correct)} = u_3 \text{(calculated)}$

iii) $u_4 \text{ (Correct)} = u_4 \text{ (calculated)} - \frac{1}{2} h^2 u_2 \text{ (calculated)} - \frac{7h^4}{240}$

Where h is the width of class intervals.

Sheppard's correction is applicable only when

1. The width of class intervals is equal.
2. The distributions are moderately skew or symmetrical.
3. The class intervals are greater than the 1/20 th of the range.
4. The total frequency is fairly large say 1000 or more.

Remarks.

Sheppard's correction is not applicable in case of J or U shaped distribution.

8.9 SELF ASSESSMENT:

1. Define moments and Explain the uses and limitations of moments.
2. What are moments? Explain the role of moments in determining the characteristics of a frequency distribution.
3. What are Sheppard's corrections?
4. From the following data calculate first four i) central moments ii) moments about 6 and iii) moments about origin.

Class intervals	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Freq.(f)	6	12	17	14	9	7	4

5. The first four moments about mean are 0, 5.507, -5.521 and 133.277. Find the first four moments about the point 6.

****

UNIT - II **LESSON-9**

SKEWNESS AND KURTOSIS

STRUCTURE

9.1 INTRODUCTION:

Earlier we have discussed various measures of central tendency and variability to understand a distribution. Sometimes there might arises a case that two distribution have same mean and standard deviation but may differ widely in their overall appearance. In both the cases mean and standard deviation is same but the two curves still differ in their overall appearance. To distinguish between such types of distributions we employ measures of Skewness and Kurtosis.

9.2 OBJECTIVES:

Skewness and Kurtosis are the terms that describe the shape and symmetry of a distribution. Skewness plays a great role in the comparison of two frequency distribution and biological investigations.

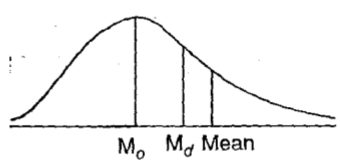
9.3 SKEWNESS

Lack of symmetry of tails (about mean) of a frequency distribution curve is known as skewness. Symmetry of tails means that the frequency of the points at equal distances on both sides of the centre of the curve on X-axis is same. Also, the area under the curve at equidistant intervals on both sides of the centre is also equal. Departure from symmetry leads to skewness. It is adjudged by the elongation of the right and left tails of the curve.

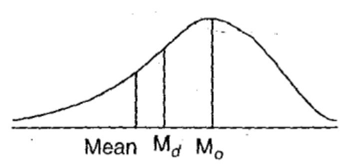
It is of two types (1) Positive Skewness (2) Negative Skewness.

(1) ***Positive Skewness:*** The Skewness is positive if the larger tail of the distribution lies towards the higher values of the variate (the right), i.e., if the curve drawn with the help of the given data is stretched more to the right than to the left

(2) ***Negative Skewness:*** The Skewness is negative if the larger tail of the distribution lies towards the lower values of the variate (left), i.e., if the curve drawn with the help of the given data is stretched more to the left than to the right.



(Positively Skewed Distribution)



(Negatively Skewed Distribution)

Q. What purpose is served by measuring Skewness?

Ans. Measure of Skewness indicates to what extent and in what direction the distribution of a variable differs from symmetry of a frequency curve. The curve may have positive or negative skewness. Both positive and negative skewness can never occur simultaneously.

Measures of Skewness. Various measures of Skewness are:

(1) Sk=M-Md, (2) Sk=M-MO, (3) Sk= (Q3—Md)—(Md—Ql).

where M is the mean, Md, the median and Mo, the mode of the distribution.

These are the absolute measures of Skewness. For comparing two series we do not calculate these measures but we calculate the relative measures called the coefficients of skewness which are pure numbers independent of units of measurement.

I. Prof Karl Pearson's Coefficient of Skewness.

$S_K = \frac{(M - M_O)}{\sigma}$ σ the standard deviation of the distribution.

If mode is ill-defined, then using the empirical relation, Mo = 3 Md — 2M, for a moderately asymmetrical distribution, we get

$$S_K = \frac{3(M - M_d)}{\sigma}$$

we observe that Sk = 0 if M = M0 = Md. Hence for a symmetrical distribution, mean, median and mode coincide.

Skewness is positive if M > M0 or M > Md and negative if M <M0 or M <Md.

Limits for Karl Pearson's Coefficient of Skewness: $S_K = \frac{3(M - M_d)}{\sigma}$

$$|M - Md| = \left|\frac{1}{n}\sum_{i=1} x_i - Md\right| = \left|\frac{1}{n}\sum_{i=1}(x_i - Md)\right| \le \frac{1}{n}\sum_{i=1}|x_i - Md| \le \frac{1}{n}\sum_{i=1}|x_i - \bar{x}| \ldots (*)$$

($\because$ The sum of the absolute deviations is minimum when taken about median.)

$$|S_k|^2 = \left|\frac{3(M - Md)}{\sigma}\right|^2 \le \frac{\left[3\cdot\frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|\right]^2}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\left[3\sum_{i=1}^{n}\left|x_i - \bar{x}\right|\right]^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \{\text{From } (*)\} \quad \ldots (**)$$

Now, Using Cauchy-Schwartz inequality:

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \le \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right), \quad \text{with } b_i = 1;\ i = 1, 2, \ldots, n, \text{ we get}$$

$$\left(\sum_{i=1}^{n} a_i\right)^2 \le \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} 1\right) \quad \Rightarrow \quad \frac{\left(\sum_{i=1}^{n} a_i\right)^2}{n\sum_{i=1}^{n} a_i^2} \le 1 \quad \ldots (***)$$

$$\therefore \quad |S_K|^2 \le 3^2 \quad [\text{From } (**) \text{ and } (***)] \quad \Rightarrow \quad |S_k| \le 3 \quad \text{or} \quad -3 \le S_k \le 3.$$

Hence, the limits for Karl Pearson's coefficient of skewness are ± 3.

2. Prof Bowley's Coefficient of Skewness. Based on quartiles: In a symmetrical distribution first quartile Q1 and Third quartile Q3 are at equal distance from Q2 , using this fact Bowley suggested a measure of skewness also called quartile coefficient of skewness.

$$S_k(B) = \frac{(Q_3 - M_d) - (M_d - Q_1)}{(Q_3 - M_d) + (M_d - Q_1)} = \frac{(Q_3 - 2M_d + Q_1)}{(Q_3 - Q_1)}$$

We see that the distribution is symmetrical if $Q_3 - M_d = M_d - Q_1$

i.e quartile are equidistant from the median(Q2).

Skewness is positive if $Q_3 - M_d > M_d - Q_1$ or $Q_3 + Q_1 > 2M_{d1}$

Skewness is negative $Q_3 - M_d > M_d - Q_1$ or $Q_3 + Q_1 < 2M_{d1}$

It is especially useful in situations where quartiles and median are used, viz.,

(i) When the mode is ill-defined and extreme observations are present in the data.

(ii) When the distribution has open end classes or unequal class intervals.

In these situations Pearson's coefficient of Skewness cannot he used

Limits for Bowley's Coefficient of Skewness. We know that for two real positive numbers a and b(i.e., a > 0 and b > 0), the modulus value of the difference (a-b) is always less than or equal to the modulus value of the sum (a + b), i.e.,

$$|a-b| \leq |a+b| \Rightarrow \left|\frac{a-b}{a+b}\right| \leq 1 \qquad \text{..........(1)}$$

We also know that (Q_3 -Md) and (Md-Q_1) are both non-negative. Thus taking a =(Q3 -Md) and b =(Md-Q1) in(1),we get

$$\left|\frac{(Q_3 - M_d) - (M_d - Q_1)}{(Q_3 - M_d) + (M_d - Q_1)}\right| \leq 1$$

$$\Rightarrow \quad |S_k\,(Bowley)| \leq 1 \qquad or \qquad -1 \leq S_k\,(Bowley) \leq +1.$$

Thus, Bowley's coefficient of skewness ranges from -I to +1.

3. **Coefficient of Skewness based on moments**:

$$S_k = \frac{\sqrt{\beta_1}\,(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)}$$

Here skewness is zero if either $\beta_1 = 0$ or $\beta_2 = -3$ but $\beta_2 = \frac{\mu_4}{\mu_2^{\;2}}$ cannot be negative. Hence for Sk=0 if andonly if $\beta_1 = 0$.

Thus for a symmetrical distribution $\beta_1 = 0$ and it is taken as a measure of skewness

$\sqrt{\beta_1} > o \qquad \Rightarrow Positive \quad skewness$

and $\sqrt{\beta_1} < o \qquad \Rightarrow Negative \quad skewness$.

Pearson's β and γ Coefficients.

Karl Pearson defined the following four coefficients, based upon the first four moments about mean

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}, \ \gamma_1 = +\sqrt{\beta_1} \qquad \text{and} \qquad \beta_2 = \frac{\mu_4}{\mu_2^2}, \qquad \gamma_2 = \beta_2 - 3$$

These co-efficients are pure numbers independent of units of measurement.

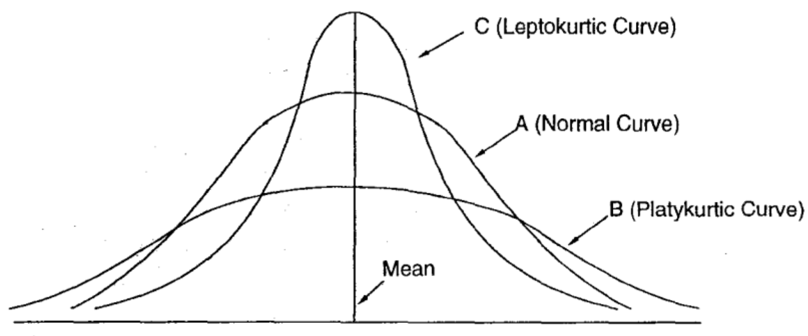
Difference between Skewness and dispersion :

1. Dispersion indicates the amount of variation rather than its direction whereas skewness indicates the direction of variation.
2. Dispersion is related to the composition of the distribution whereas skewness is related to its shape.
3. The measures of skewness are dependent upon the amount of dispersion.

9.4 KURTOSIS:

Kurtosis enables to have an idea about the flatness or peaked ness of the curve. Curve which is neither flat nor peaked is called normal curve (Mesokurtic). On the basis of degree of peaked ness or flatness Karl Pearson in 1950 introduced three terms.

(i) Mesokurtic (ii) Leptokurtic (ii) Plátykurtic



Kurtosis of any distribution is compared with the normal distribution, Curve which is neither flat nor peaked is called normal curve (Mesokurtic).

If the concentration of the items is much at the centre as compared with the normal curve. Such a distribution is called leptokurtic or distribution with +ve kurtosis.

If the concentration of the items is low at the centre as compared with the normal curve. Then its frequency curve will be flatter than normal curve. Such a distribution is called plátykurtic.

The frequency curve for a platykurtic distribution is relatively flat- topped, and for a leptokurtic distribution it has relatively high peak. A mesokurtic distribution(e.g. Normal distribution) is of moderate peakedness.

Measure of Kurtosis:

The distributions with same mean and standard deviation may be equally skewed but may differ as regards the peak of their curves.

Prof. Karl Pearson gave the following formula based upon 4th and 2nd central moment.

$$\beta_2 = \frac{\mu_4}{\mu_2^{\,2}} \quad \text{and} \quad \gamma_2 = \beta_2 - 3$$

If $\beta_2 = 3 \; or \quad \gamma_2 = 0 \Rightarrow \textit{the curve is mesokurtic}$

$\beta_2 > 3$ or $\gamma_2 > 0 \Rightarrow$ *the curve is leptokurtic*

$\beta_2 < 3$ or $\gamma_2 < 0 \Rightarrow$ *the curve is platykurtic*

Example: The first four moments of a distribution about the value 4 of the variable are -15, 17, -30 and 108. Find the moments about mean, Skewness and Kurtosis.

Find also the moments about (i) the origin, and (ii) the point X= 2.

Solution. In the usual notations, we are given

$$A = 4, \mu_1' = -1{\cdot}5, \quad \mu_2' = 17, \quad \mu_3' = -30 \text{ and } \quad \mu_4' = 108.$$

Moments about mean :

$$\mu_2 = \mu_2' - \mu_1'^2 = 17 - (-1{\cdot}5)^2 = 17 - 2{\cdot}25 = 14{\cdot}75$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = -30 - 3 \times (17) \times (-1{\cdot}5) + 2(-1{\cdot}5)^3 = 39{\cdot}75$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$
$$= 108 - 4(-30)(-1{\cdot}5) + 6(17)(-1{\cdot}5)^2 - 3(-1{\cdot}5)^4 = 142{\cdot}3125$$

Hence $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(39{\cdot}75)^2}{(14{\cdot}75)^3} = 0{\cdot}4926, \qquad \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{(142{\cdot}3125)}{(14{\cdot}75)^2} = 0{\cdot}6543$

Also Mean, $(\bar{x}) = A + \mu_1' = 4 + (-1{\cdot}5) = 2{\cdot}5$

Moments about origin : We have

$$\bar{x} = 2{\cdot}5, \quad \mu_2 = 14{\cdot}75, \; \mu_3 = 39{\cdot}75 \quad \text{and} \quad \mu_4 = 142{\cdot}31 \text{ (approx).}$$

We know : $\bar{x} = A + \mu_1'$, where μ_1' is the first moment about the point $x = A$.

Taking $A = 0$, we get the first moment about origin as $\mu_1' =$ mean, $\bar{x} = 2{\cdot}5$.

Using relation between moments about mean in terms of moments about any point

$$\mu_2' = \mu_2 + \mu_1'^2 = 14{\cdot}75 + (2{\cdot}5)^2 = 14{\cdot}75 + 6{\cdot}25 = 21$$

$$\mu_3' = \mu_3 + 3\mu_2\mu_1' + \mu_1'^3 = 39{\cdot}75 + 3(14{\cdot}75)(2{\cdot}5) + (2{\cdot}5)^3 = 166$$

$$\mu_4' = \mu_4 + 4\mu_3\mu_1' + 6\mu_2\mu_1'^2 + \mu_1'^4$$
$$= 142{\cdot}3125 + 4(39{\cdot}75)(2{\cdot}5) + 6(14{\cdot}75)(2{\cdot}5)^2 + (2{\cdot}5)^4 = 1{,}132.$$

Moments about the point $x = 2$. We have $\bar{x} = A + \mu_1'$. Taking $A = 2$, the first moment about the point $x = 2$ is $\mu_1' = \bar{x} - 2 = 2{\cdot}5 - 2 = 0{\cdot}5$.

Hence $\mu_2' = \mu_2 + \mu_1'^2 = 14{\cdot}75 + 0{\cdot}25 = 15$

Example: Show that for discrete distribution $\beta_2 > 1$.

Solution. We have to show that $\mu_4 / \mu_2^2 > 1$,

i.e, $\mu_4 > \mu_2^2$

If x_i / f_i is a frequency distribution having n observations $X_1, X_2 ....X_n$

with corresponding frequencies $f_1, f_2, f_3 ....f_n$ then we have to prove that

$\mu_4 > \mu_2^2$, *i.e.,* $\frac{1}{N}\sum_i f_i (x_i - \bar{x})^4 > \left[\frac{1}{N}\sum_i f_i (x_i - \bar{x})^2\right]^2$ Putting $(x_i - \bar{x})^2 = z_i$ we

have to show that

$$\frac{1}{N}\sum_i f_i z_i^2 > \left(\frac{1}{N}\sum_i f_i z_i\right)^2 \Rightarrow \frac{1}{N}\sum_i f_i z_i^2 - \left(\frac{1}{N}\sum_i f_i z_i\right)^2 > 0 \text{ or } \sigma_z^2 > 0,$$

This is always true, since variance is always positive.

Hence $\beta_2 > 1$

Example: Calculate the measure of skewness based on quartiles and median from the following data:

Variables	Frequency
10-20	358
20-30	2417
30-40	976
40-50	129
50-60	62
60-70	18
70-80	10

Ans. Skewness based on quartiles and median.

$$\text{Coeff. Of Sk.} = \frac{Q_3 + Q_1 - 2Med.}{Q_3 - Q_1}$$

Calculation of Q_1, Q_3 and median

C. I	f	C.f.
10-20	358	358
20-30	2417	2775
30-40	976	3751
40-50	129	3880
50-60	62	3942
60-70	18	3960
70-80	10	3970

Q_1 = size of $\frac{N}{4}$ th items

= size of $\frac{3970}{4}$ = 992.5 th item.

Q_1 lies in the class 20-30

$$Q_1 = L + \frac{N/4 - c.f}{f} \times i \quad = 20 + \frac{634.5}{2417} \times 10$$

=20+2.66 = 22.66

Q_3 = size of $\frac{3N}{4}$ th items

=size of 2977.5 th item

Q_3 lies in the class 30-40

$$Q_3 = L + \frac{3N/4 - c.f}{f} \times i \quad = 30+2.26 \quad =32.26$$

Med. = size of $\frac{N}{2}$ th items

= size of 1985 th item

Median lies in the class 20-30

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

Med. = 20+6.73 =26.73

Coeff. Of Sk. $= \frac{Q_3 + Q_1 - 2Med.}{Q_3 - Q_1} = \frac{32.26 + 22.66 - 2(26.73)}{32.26 - 22.66}$

Coeff. Of Sk = .152

Example. Show that for a discrete distribution, $\beta_2 > \beta_1$

Solution: Here

$$\beta_2 = \frac{\mu_4}{u_2{}^2} \qquad \beta_1 = \frac{u_3{}^2}{u_2{}^3}$$

$\therefore \quad \beta_2 > \beta_1$

If

$\frac{\mu_4}{u_2{}^2} > \frac{u_3{}^2}{u_2{}^3}$ or if $u_2 u_4 > u_3{}^2$

Or if $\frac{1}{N}\sum f_i (x_i - \bar{x})^2 . \frac{1}{N}\sum f_i (x_i - \bar{x})^4 > \left[\frac{1}{N}\sum f_i (x - \bar{x})^3\right]^2$

0r if $\sum f_i y_i{}^2 \sum f_i y_i{}^4 > \left(\sum f_i y_i{}^3\right)^2$ where $y_i = x_i - \bar{x}$

Or if $\left(f_1 y_1{}^2 + f_2 y_2^2 + ...\right)\left(f_1 y_1{}^4 + f_2 y_2^4 + ...\right) > \left(f_1 y_1{}^3 + f_2 y_2^3 + ...\right)^2$

Or if

$\left(f_1^2 y_1{}^6 + f_2{}^2 y_2^6 + f_1 f_2 y_1^2 y_2^2 .(y_1^2 + y_2^2.)\right) + ... > (f_1^2 y_1{}^6 + f_2^2 y_2^6 + 2 f_1 . f_2 y_1^3 y_2^3 + ...)$

Or if $f_1 f_2 y_1^2 y_2^2 (y_1^2 + y_2^2) - 2 f_1 f_2 y_1^3 y_2^3 + ... \succ 0$

Or if $f_1 f_2 y_1^2 y_2^2 (y_1^2 + y_2^2 - 2 y_1 y_2 + ... \succ 0$

Or if $f_1 f_2 y_1^2 y_2^2 (y_1 - y_2)^2 + ... \succ 0$

Since f_i (i=1,2,….n) is always +ve, the inequality is always true.

hence $\beta_2 > \beta_1$

Example. Calculate the first four moments about mean for the following data.

X:	1	2	3	4	5	6	7	8	9
F:	1	6	13	25	30	22	9	5	2

Calculate β_1 and β_2

Solution: Calculation for moments

X	f	d= x-5	fd	fd^2	fd^3	fd^4
1	1	-4	-4	16	-64	256
2	6	-3	-18	54	-162	486
3	13	-2	-26	52	-104	208
4	25	-1	-25	25	-25	25
5	30	0	0	0	0	0
6	22	1	22	22	22	22
7	9	2	18	36	72	144
8	5	3	15	45	135	405
9	2	4	8	32	128	512
	113	0	-10	282	2	2058

Moments about the point x=5

d = x-A = x-5 (A=5) N =11

$$u_1' = \frac{\sum fd}{N} = \frac{-10}{113} = -.0885$$

$$u_2' = \frac{\sum fd^2}{N} = \frac{282}{113} = 2,4956$$

$$u_3' = \frac{\sum fd^3}{N} = \frac{2}{113} = .0177$$

$$u_4' = \frac{\sum fd^4}{N} = \frac{2058}{113} = 18.2124$$

Moments about mean

u_1 = First moment about mean is always equal to zero

$$u_2 = u_2' - u_1'^2 = 2.4878$$

$$u_3 = u_3' - 3u_2'u' + 2u_1'^3 = .0177\text{-}3(2.,4956)(\text{-}.0885)\text{+}2\,(-.0885)^3$$

$$u_3 = .6789$$

$$u_4 = u_4' - 4u_3'u_1' + 6u_2'(u_1')^2 - 3(u_1')^4$$

After simplification we get

$$u_4 = 18.3357$$

Hence $\beta_1 = \dfrac{{u_3}^2}{{u_2}^3} = \dfrac{(.6789)^2}{(2.4878)^3} = \dfrac{.4609}{15.3974} = .0299$

$$\beta_2 = \frac{\mu_4}{{u_2}^2} = \frac{18.3357}{6.1891} = 2.9626$$

Hence we conclude slightly symmetrical distribution since β_1 =.0299

From β_2 we conclude that distribution is platykurtic since its values less than 3.

9.5 SELF ASSESSMENT QUESTION:

1) Explain the term Skewness and Kurtosis used in connection with the frequency distribution of a continuous variable. Give the measure of skewness and kurtosis.

2) Calculate coefficients of Skewness from the following data:

Marks above:	0	10	20	30	40	50	60	70	80
Frequency:	150	140	100	80	80	70	30	14	0

3) Calculate the coefficient of Kurtosis from the following data:

Variate:	1	2	3	4	5	6	7	8	9
Frequency:	1	6	13	25	30	22	9	5	2

4) The fourth moment about mean of a frequency distribution is 48. what must be the value of its standard deviation in order that the distribution be

1) Leptokurtic 2) Mesokurtic 3) Platykurtic

5) How can you find out Skewness and Kurtosis of a distribution from moment about the mean? Show that for any distribution $\beta_2 > 1$.

****

UNIT - III **LESSON - 10**

BI VARIATE DATA, SCATTER DIAGRAM AND CORRELATION

STRUCTURE:

10.1 INTRODUCTION

So far we have studied some characteristics of one variable only, e.g. mean of distribution of height, standard deviation of weight, skewness of the distribution of income. But many situations arise in which we may have to study two variables simultaneously, say x and y and we may be interested to measure numerically the strength of this association between the variables. Such a measure will determine how well a linear or other equation explains the relationship between two variables. Further, if we measure more than two variables on each unit of a distribution, it is called a multivariate distribution. On the other hand from scatter diagram we can form a fairly good, though rough, idea about the relationship between the two variables.

10.2 OBJECTIVES:

With the help of bivariate distribution, we may be interested to find if there is any relationship between the two variables under study. The study of relationship between two or more variables is most important problems in life. It is a common knowledge that many of variables show some relationship or the other. The statisticians have developed methods of expressing such relationship in number.

10.3 BIVARIATE DATA:

In many situations we came across that two factors or characters are used for the study under consideration at a time .These factors or characters may have different units of measurements and moreover the range or the domain of the data may also be different. Such types of data are known as bivariate data. For example, if we measure the height and weight of ten persons. We obtain two series. One for the height and other for the weight. Such series forms the bivariate data. There may be examples of income and expenditure of certain locality, the marks obtained in Economics and Statistics of a class in an examination. One of the examples of bivariate data is the following series given as under:-

Marks in Economics: 25, 38, 42, 47, 35, 37, 60, 59, 42, 49

Marks in Statistics: 21, 24, 29, 37, 51, 38, 55, 50, 47, 51

are the example of bivariate data.

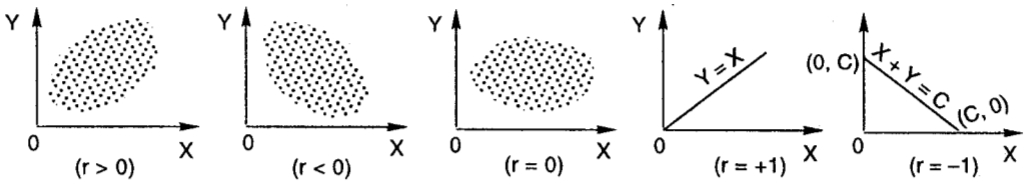
In a bivariate data, we may be interested to find out if there is any relationship or co variations between the variables under study exist or not.

10.4 SCATTER DIAGRAM :

It is the simplest method of studying relationship. Under this method x variable is measured on the horizontal axis and y variable on the vertical axis and for each pair of x and y values, points are plotted.

Scatter diagram is the simplest way of the diagrammatic representation of the bi variate data. It gives us notion about the presence or absence of correlation. When the pairs of values (, y) for i=1, 2, 3………,n are plotted on a graph paper, the point shows the pattern in which they lie. Such a diagram is known as scatter diagram. If these points on a straight line, it expected that there is a linear relationship between X and Y, otherwise not. From the scatter diagram we can form fairly idea whether the variables are correlated or not. For example, if the points or dots are very close to each other, we can expect a good amount of correlation between the variables and if the dots are widely scattered, we expect a poor amount of correlation between them.

The following diagrams of the scattered data depict different forms of correlation.



In the above figure (a) both the variables rises with the constant ratio and the point's forms straight line. Thus there is a positive correlation.

In the above figure (b), the dots show a downward trend .So, there is a negative correlation.

In figure (c) the points are widely scattered, which is impossible to see any form of trend. Thus there is no correlation or zero correlation

In the figure (d) the points are fall in same ratio , rising trend is called perfect positive correlation.

In the figure (e) one variable rises and the other falls in the same ratio .Thus points form a straight line have a perfect negative correlation.

Scatter diagram is a simple and attractive method of ascertaining the nature of correlation between two variables. At a glance, one can know whether variables are correlated or not, and, if they are correlated whether correlation is positive or negative. Scatter diagram can usually give the following information about the variables being correlated.

1) It tells how closely the two variables under study are related. This is useful information which indicates how accurately we can estimate values of the dependent variable on the basis of the independent variable.

2) It indicates the direction of correlation, i.e. whether the variables move in the same direction(positive correlation) or in the opposite direction(negative correlation)

3) The shape of scatter of points tells whether the correlation is linear or curvilinear.

4) The scatter diagram may be used to estimate values of the dependent variable when the values of the independent variable are given.

5) The scatter diagram may also be used in the detecting of abnormal points.

6) It helps in obtaining the line of best fit.

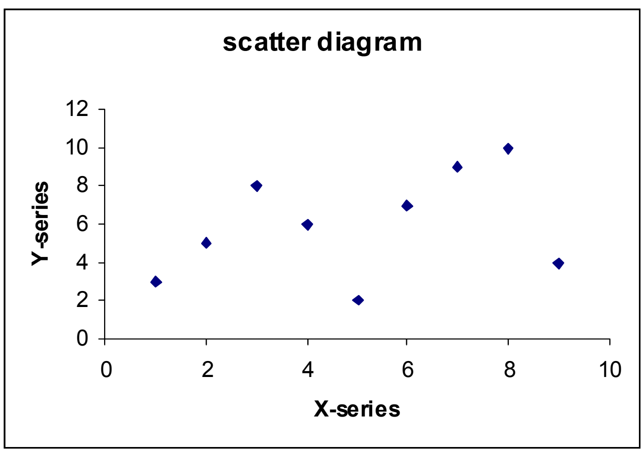
One of the serious limitations of scatter diagram is that the degree of correlation cannot be known by this method. It gives only a rough idea of how the two variables are related, but definite conclusions cannot be drawn merely examining the diagram.

Examples (solved and unsolved)

Example1: From the following pairs of values of the variable X and Y, draw a scatter diagram and interpret the result.

X	1	2	3	4	5	6	7	8	9
Y	3	5	8	6	2	7	9	10	4

Solution: Scatter Diagram



From the scatter diagram we observe that correlation between X and Y moderately high degree.

10.5 CORRELATION:

When two phenomenon's are varying simultaneously in the same direction or in the opposite direction and the variation in the one variable are followed by the variation in the other, the two phenomenons are said to be correlated.

In the words of W.I. King, "correlation means that between two series or group of data there exist some causal relationship."

According to prof. Boddington,"whenever some definite connection exists between two or more groups, classes or series of data there is said to be correlation."

In statistical analysis, the concept and technique of correlation is of great importance. In 1896, Karl Pearson introduced a method of ascending correlation by means of coefficient of correlation. Measuring of correlation ensures that prediction of variable will be reliable and near to the reality.

Product Moment Correlation Coefficient

Correlation is the measure of the degree of relationship between two variables. For example, there exist some relationship between the income and expenditure of an individual, Demand & supply, height & weight, production & rainfall etc. correlation is a linear relationship between two variables which measures the degree of their intensity.

If the change in one variable affects the change in other variable, the variables are said to be correlated. It is denoted by r(x, y). Karl Pearson developed a formula called correlation coefficient between two random variables X & Y , is a mathematical measure of linear relationship between them and is defined as

$$r(X, Y) = \frac{COV(X,Y)}{\sigma_X \sigma_Y}$$

If (x_i, y_i); i = 1, 2, ………, n, is the bivariate distribution, then

$$\text{Cov}(X, Y) = E[\{X - E(X)\}\{(Y - E(Y)\}] = \frac{1}{n} \Sigma (x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{\frac{1}{n} \Sigma (x_i - \bar{x})(y_i - \bar{y})}{\left[\frac{1}{n} \Sigma (x_i - \bar{x})^2 . \frac{1}{n} \Sigma (y_i - \bar{y})^2\right]^{1/2}}$$

$$\sigma_X^2 = E\{X - E(X)\}^2 = \frac{1}{n} \Sigma(x_i - \bar{x})^2 \quad \text{and} \quad \sigma_Y^2 = E\{Y - E(Y)\}^2 = \frac{1}{n} \Sigma(y_i - \bar{y})^2$$

$$r(X, Y) = \frac{COV(X,Y)}{\sigma_X \sigma_Y} =$$

Another convenient form of the formula for computational work is as follows

$$\text{Cov}(X, Y) = \frac{1}{n}\sum x_i y_i - \bar{x}\,\bar{y}, \quad \sigma_X^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2 \text{ and } \sigma_Y^2 = \frac{1}{n}\sum y_i^2 - \bar{y}^2$$

It is a mathematical fact that the value of correlation coefficient (r) always lie between -1 and +1. the extreme value of -1 and +1 are obtain only situations where there is perfect linear relationship between the two variables exists.

When r = -1, which means that the relationship is perfectly negative and

r = +1, indicates that the relationship is perfectly positive.

Also r = 0, means that there is no relationship between the variable X & Y.

Assumption of Karl Pearson's Correlation Coefficient:-

Pearsonian correlation coefficient r is based on the following assumptions:-

1) The variables X & Y under study are linearly related. In other words, the scatter diagram of the data will give a straight line curve.
2) Each of the correlated bivariate data is being affected by a large number of independent causes is operating so as to produce normal distribution.
3) The forces so operating on each of the variable series are not independent of each other but are related in a causal fashion. In other words, cause and affect relationship exist between different forces operating on the items of the two variable series.

Types of correlations

On the basis of nature of relationship between the variables, i.e. the direction in which changes takes place in them or the ratio by which they change, correlation may be;-

1) Positive or Negative.
2) Simple, Partial or Multiple.
3) Linear or Non Linear.

1) Positive or Negative Correlation :-

Positive or direct correlation refers to the movement of the variables in the same direction. as one variable increases the other also increases and as one variable decreases, the other also decreases. These sorts of correlation exist between supply & price of the commodity For example, when the price of the commodity increases, the supply of that commodity also increases or vice-versa.

'Negative' or 'inverse' correlation refers to, where one variable increases or decreases the other moves in the reverse direction. Such a correlation is found between price & demand. When price of a commodity increases its demand decreases or vice versa.

Positive correlation				Negative correlation			
(a) both variable increases		b) both variable decreases		(c) one variable increasing other decreasing		d) one variable decreasing other increasing	
X	Y	X	Y	X	Y	X	Y
70	35	27	44	60	42	27	25
75	40	23	40	65	38	23	32
80	45	20	38	72	33	20	36
85	50	19	36	77	29	19	39
90	55	17	32	82	25	17	42
95	60	13	30	86	18	13	44

2) Linear or Non Linear Correlation: -

The distinction between linear or non linear correlation is based upon the consistency of the ratio of change between the variable under study. If the ratio of change between two variable is uniform then there will be linear correlation between them. Their relationship is best described by a straight line. In non linear relationship, which can also be said curvilinear, the amount of change in one variable does not bear a constant ratio to the amount of change in the other variables. The graph of such variables having such a relationship will form a curve.

The following data show this phenomenon:

Linear correlation		Non-Linear correlation	
X	Y	X	Y
10	40	40	10
30	90	45	12
50	140	60	22

70	190	90	24
90	250	98	45
100	300	120	56

We can find curvilinear relationship in the most of the phenomena. However, since the technique of non-linear correlation analysis being very complicated one, we generally assume the relationship between the variables under study is linear.

Degree of relationship:

The intensity of relationship between two variables can be ascertained by the quantitative value of coefficient of correlation which can be found out by calculation, it can be determined as given below :

1) Perfect correlation
2) Absence correlation
3) Limited degree correlation

Karl person has given a formula for measuring correlation .The result of this formula(r) varies between.

Degree of correlation	Positive	Negative
Perfect correlation	+1	-1
Very high degree of correlation	+.9 or more	-.9 or more
Sufficiently high degree of correlation	From +.75 to +.9	From -.75 to -.9
Moderate degree of correlation	From +.6 to +.75	From -.6 to -.75
Only the possibility of a correlation	From +.3 to +.6	From -.3 to -.6
Possibly no correlation	Less than +.3	Less than -.3
Absence of correlation	0	0

Example: Below are given the marks of 15 students' mathematics and statistics, the maximum marks being 50 in each paper. Calculate karl Pearson,s co-efficient of correlation and stat what you infer from it.

Roll no.:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
MathX:	26	26	27	28	28	29	30	25	19	21	22	23	23	24	24
StatY:	17	20	21	20	22	22	23	18	14	16	15	14	17	14	17

Ans : computation of co-efficient of correlation.

$$r(X, Y)= \frac{COV(X,Y)}{\sigma_X \sigma_Y}$$

Roll no.	(X)	(Y)	X-25(x)	Y-18(y)	x^2	y^2	xy
1	26	17	1	-1	1	1	-1
2	26	20	1	2	1	4	2
3	27	21	2	3	4	9	6
4	28	20	3	2	9	4	6
5	28	22	3	4	9	16	12
6	29	22	4	4	16	16	16
7	30	23	5	5	25	25	25
8	25	18	0	0	0	0	0
9	19	14	-6	-4	36	16	24
10	21	16	-4	-2	16	4	8
11	22	15	-3	-3	9	9	9

12	23	14	-2	-4	4	16	8
13	23	17	-2	-1	4	1	2
14	24	14	-1	-4	1	16	4
15	24	17	-1	-1	1	1	1

$\sum X = 375 \quad \sum Y = 270$

$\sum x^2 = 136 \; \sum y^2 = 138 \; \sum xy = 122$

$$r = \frac{122}{\sqrt{136 \times 138}} == 0.891$$

The coefficient of correlation is .891 which indicates there is a high degree of positive correlation between marks in mathematics and statistics.

10.6 SUMMARY:

In this lesson we conclude that scatter diagram only tells us about the nature of the relationship whether it is positive or negative and whether it is high or low. It does not provide us an exact relationship. But in simple correlation, we have to established mathematical relationship between two variables

10.7 SELF ASSESSMENT QUESTION:

1. What is a scatter diagram? Explain how this can be used to indicate the degree and type of association between two variables.
2. Construct the scatter diagram of the data given below. Draw a free hand straight line through the group of points and from your diagram and disuss the probable amount of correlation.

Year:	1980	1981	1982	1983	1984	1985	1986	1987
Cotton import:	47	64	100	97	126	203	172	115
Cotton export:	70	85	100	103	111	139	133	115

3. Compute karl pearson correlation coefficient between X and Y from the following data.

X:	1	2	3	4
Y:	1	4	9	16

Is the correlation coefficient equal to unity? If not, why not?

4. Define correlation. Explain various types of correlation with suitable examples.

5. Distinguish clearly between:

   i) Positive and negative correlation

   ii) Linear and non-linear correlation

****

UNIT-III **LESSON - 11**

PRODUCT MOMENT CORRELATION COFFECICENT & ITS PROPERTIES

STRUCTURE:

11.1 INTRODUCTION:

In this lesson, we will study problems involving calculations of correlation. There are many situations in which we are confronted with paired measurements & must examine whether a relationship exits between the two variables under consideration, we may also like to know the strength of that relationship.

Correlation refers to the relationships of variables. Some, relationship is found in certain types of variables. For example, there exits a relationship between price and demand, production and employment, wages and price index etc. Even everyday experience demonstrates, how far different phenomenons are related to each other in some way. For example, it is found that a close relationship between age of Husbands and Wives, capital invested and profit earned etc. correlation is a statistical technique which analysis the degree or extent to which two variables or phenomenon fluctuate with reference to each other. Correlation denotes the interdependence between two variates. Croxton & Cowden states that, "when the relationship is of quantitative nature, the appropriate statistical tools for discovering and measuring the relationship and expressing it in a brief formula known as Correlation."

11.2 OBJECTIVES:

On going through the lesson you will be enable to:-

* Compute the strength of correlation coefficient between two variables.
* Determine whether you analyses that the variables are correlated.
* In 1896, Karl Pearson introduced a method of ascending correlation by means of coefficient of correlation. Measuring of correlation ensures that prediction of variable will be reliable and near to the reality.

11.3 LINEAR AND NON-LINEAR CORRELATION:

The correlation between two variables is said to be linear if corresponding to a unit change in one variable, there is a constant change in the other variable over the entire range of the values. If the amount of change in one variable tends to bear a constant ratio to the amount of change in the other variable, then the correlation is said to be linear. Thus one finds, ordinarily, that if income is doubled, wealth is doubled or if wealth is doubled income is doubled. If the amount of change in one variable does not bear a constant ratio to the amount of change in the other variable, then correlation is said to be non-linear or curvilinear. Thus if one should increase the amount of rainfall the amount of wheat per acre is not necessarily doubled. It should be noted that the cases of linear relationship are rare.

11.4 PRODUCT MOMENT CORRELATION COEFFICIENT:

Correlation is the measure of the degree of relationship between two variables. For example, there exist some relationship between the income and expenditure of an individual, Demand & supply, height & weight, production & rainfall etc. correlation is a linear relationship between two variables which measures the degree of their intensity.

If the change in one variable affects the change in other variable, the variables are said to be correlated. It is denoted by p r(x, y). Karl Pearson developed a formula called correlation coefficient between two random variables X & Y , is a mathematical measure of linear relationship between them and is defined as :-

$$r(x, y) = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

If (x, y); i = 1,2,3........, n is the biavariate distribution , then

Cov (x, y) =E [{x - E(x)} {y - E(y)}]

$$\sigma_{x^2} = E[\{x - E(x)\}]^2$$

$$\sigma_{y^2} = E[\{Y - E(Y)\}]^2$$

Assumption of Karl Pearson's Correlation Coefficient:-

Pearsonian correlation coefficient r is based on the following assumptions:-

1) The variables X & Y under study are linearly related. In other words, the scatter diagram of the data will give a straight line curve.

2) Each of the correlated bivariate data is being affected by a large number of independent causes is operating so as to produce normal distribution.

3) The forces so operating on each of the variable series are not independent of each other but are related in a causal fashion. In other words, cause and affect relationship exist between different forces operating on the items of the two variable series.

Properties of Product Moment correlation coefficient:

1. **To** prove: $-1 \le r \le 1$

Proof. Let us consider the sum of squares $\sum\left[\frac{x-\bar{x}}{\sigma_x} \pm \frac{y-\bar{y}}{\sigma_y}\right]^2$, which is always non – negative.

i.e. $$\sum\left[\frac{x-\bar{x}}{\sigma_x} \pm \frac{y-\bar{y}}{\sigma_y}\right]^2 \ge 0$$

$$= \sum\left[\left(\frac{x-\bar{x}}{\sigma_x}\right)^2 + \left(\frac{y-\bar{y}}{\sigma_y}\right)^2 \pm 2\left(\frac{x-\bar{x}}{\sigma_x}\right)\left(\frac{y-\bar{y}}{\sigma_y}\right)\right] \ge 0$$

$$\Rightarrow \frac{\sum(x-\bar{x})^2}{\sigma_{x^2}} + \frac{\sum(y-\bar{y})^2}{\sigma_{y^2}} \pm \frac{2\sum(x-\bar{x})(y-\bar{y})}{\sigma_x \sigma_y} \ge 0$$

Dividing throughout by n, the number of pairs observation, we get

$$\frac{1}{\sigma_{x^2}}\frac{1}{n}\sum(x-\bar{x})^2+\frac{1}{\sigma_{y^2}}\frac{1}{n}\sum(y-\bar{y})^2 \pm \frac{2\sum(x-\bar{x})(y-\bar{y})}{\sigma_x\sigma_y n} \geq 0$$

$$\Rightarrow \quad \frac{1}{\sigma_{x^2}}\sigma_{x^2}+\frac{1}{\sigma_{y^2}}\sigma_{y^2} \pm \frac{2Cov(x,y)}{\sigma_x\sigma_y} \geq 0$$

$$\Rightarrow \quad 1+1 \pm 2r \geq 0$$

Hence $-1 \leq r \leq 1$

Remarks: This theorem provides us a check on our calculations.

2. Show that coefficient of correlation is independent of change of origin & scale.

Proof:

Let us consider $u=\frac{x-a}{h}$ and $v=\frac{y-b}{k}: h,k \succ 0$

Where a, b, h and k are constants then correlation coefficient between x and y is same as correlation coefficient between u and v i.e.,

r(x,y) = r(u,v) $\Rightarrow r_{xy}=r_{uv}$

From the above we have

X = a+hu and y = b+kv ……………….*

Summing both sides and dividing by n, we get

$\bar{x}=a+h\bar{u}$ and $\bar{y}=b+k\bar{v}$ ……………………………**

Subtracting (**) from (*) we get

$x-\bar{x}=h(u-\bar{u})$ and $y-\bar{y}=k(v-\bar{v})$

Now $$r_{xy}=\frac{\sum\left[h(u-\bar{u})k(v-\bar{v})\right]}{\sqrt{\sum h^2(u-u)^2\sum k^2(v-\bar{v})^2}}=\frac{hk\sum(u-\bar{u})(v-\bar{v})}{hk\sqrt{\sum(u-\bar{u})^2\sum(v-\bar{v})^2}}$$

$$=\frac{\sum(u-\bar{u})(v-\bar{v})}{\sqrt{\sum(u-\bar{u})^2\sum(v-\bar{v})^2}}=r_{uv}$$, as desired

This is one of the important properties of the correlation coefficient and is extremely helpful in numerical computation of r.

3. Two independent variable are uncorrelated but the converse is not true.

Proof. We know that

Cov(x,y) =E(xy)-E(x)E(y)………*

Since the expected value of the variable is nothing but its arithmetic mean.

If x and y are independent variables then

E(xy) = E(x)E(y)

Substituting in *, we get

Cov(x,y) = E(x)E(y)- E(x)E(y)=0

Hence if x and y are independent variables then

$$r_{xy} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = 0$$

$\Rightarrow$ independent variables are un correlated.

However, the converse of the theorem is not true i.e. , un correlated variables need not necessarily be independent.

4) Show by an example that two uncorrelated variables may not be independent.

Proof: Let X & Y be two variables which takes the values

X	-3	-2	-1	1	2	3	$\sum x = 0$
Y	9	4	1	1	4	9	$\sum y = 28$
XY	-27	-8	-1	1	8	27	$\sum xy = 0$

$$r_{xy} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = 0$$

Thus in the above example, the variable X & Y are uncorrelated. But on careful examination we find that X & Y are not independent, they are connected by the relation Y= X^2.

Hence two uncorrelated variables need not necessarily be independent. A simple reasoning for this strange conclusion is that r(X,Y)=0 merely implies the absence of any linear relationship between the variables X & Y .There may , however, exist some other form of relationship between them e.g. quadratic, cubic or trigonometric, etc.

5. If **each** of two variables X and Y takes values ,0 ,1 with probabilities, then r(X,Y)=0 this implies that X and Y are independent.

Proof:

Let X takes the values 1 and 0 with positive probabilities p_1 and q_1 respectively and let Y take the values 1 and 0 with positive probabilities p_2 and q_2 respectively. Then

r(X,Y) =0 $\Rightarrow$ Cov(X,Y) = 0

$\Rightarrow$ 0 = E(XY)-E(X)E(Y) =1.P(X=1 $\cap Y$ =1)-P(X=1) $\times$ P(Y=1)

0 =P(X=1 $\cap Y$ =1)- $P_1 P_2$

$\Rightarrow$ P(X=1 $\cap Y$ =1) = $P_1 P_2$ =P(X=1).P(Y=1)

Hence X and Y are independent

Example: The variables X and Y are connected by the equation aX + bY + c=0. Show that the correlation between them is -1 if signs of a & b are alike & +1 if they are different.

Solution:

The equation is ax + bY + c=0 ……………(1)

Taking expectation on both sides we have

a E(X) + b E(Y) + c= 0 ……………(2)

Subtract (2) from (1) we have

a[X-E(X)]+b[Y-E(Y)]=0

$$X-E(X) = -\frac{b}{a}\,[Y-E(Y)]$$

Cov (X, Y) = E [{X-E(X)} {Y-E(Y)}]

$$= -\frac{b}{a} E(Y-E(Y))^2 = \frac{-b}{a}\sigma_{y^2}$$

$$\sigma_{x^2} = E\left(X - E(X)\right)^2 = \frac{b^2}{a^2} E(Y - E(Y))^2 = \frac{b^2}{a^2}\sigma_{y^2}$$

$$\therefore \quad r_{xy} = \frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y} = \quad - \frac{b/a \, \sigma_{y^2}}{\sqrt{b^2/a^2}\,\sigma_{y^2}} = \frac{-b/a\,\sigma_{y^2}}{\left|\frac{b}{a}\right|\sigma_{y^2}} = \frac{-b/a}{\left|\frac{b}{a}\right|}$$

r_{xy} = +1, if a & b are of opposite signs

= -1, if a & b are of same signs

Example: If X,Y are standardized random variables and

$$r(aX+bY,bX+aY) = \frac{1+2ab}{a^2+b^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..*$$

find r (X,Y), the coefficient of correlation between X and Y.

Proof: Since X and Y are standardized random variables, E(X) = E(Y)=0

And var(X)=var(Y)=1 $\Rightarrow$ E(X^2)= $E(Y^2)$ =1

Cov(X,Y)=E(XY) $\Rightarrow$ E(XY)=r(X,Y) $\sigma_x \sigma_y$ =r(X,Y)

Also we have

r(aX+bY,bX+aY)

$$= \frac{E\left[(aX+bY)(bX+aY)\right] - E(aX+bY)E(bX+aY)}{\left\{\operatorname{var}(aX+bY).\operatorname{var}(bX+aY)\right\}^{\frac{1}{2}}}$$

$$= \frac{E\left[abX^2 + a^2XY + b^2YX + abY^2\right] - 0}{\left\{\left[a^2V(X) + b^2V(Y) + 2abCov(X,Y)\right]\left[b^2V(X) + a^2V(Y) + 2abCov(X,Y)\right]\right\}^{\frac{1}{2}}}$$

$$= \frac{a.b.1 + a^2r(X,Y) + b_2r(X,Y) + ab.1}{\left\{a^2 + b^2 + 2abr(X,Y)\right\}^{\frac{1}{2}}\left\{b^2 + a^2 + 2abr(X,Y)\right\}^{\frac{1}{2}}}$$

$$= \frac{2ab + (a^2+b^2).r(X,Y)}{a^2+b^2+2ab.r(X,y)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..**$$

from (*) and(**), we get

$$\frac{1+2ab}{a^2+b^2} = \frac{(a^2+b^2).r(X,Y)+2ab}{a^2+b^2+2ab.r(X,Y)}$$

Cross multiplying, we get

$$(a^2+b^2)(1+2ab)+2ab.r(X,Y)(1+2ab)=(a_2+b^2)^2.r(X,Y)+2ab(a^2+b^2)$$

$$\Rightarrow (a^4+b^4+2a^2b^2-2ab-4a^2b^2).r(X,Y)=(a^2+b^2)$$

$$\Rightarrow \left[(a^2-b^2)^2-2ab\right]r(X,Y)=a^2+b^2$$

(

$$\Rightarrow r(X,Y)=\frac{a^2+b^2}{(a^2-b^2)^2-2ab}$$

Example: the random variables X and Y are jointly distributed and U and V are defined by

U=Xcosα+Ysinα,

V=Ycosα-Xsinα

Show that U and V will be uncorrelated if

$$\tan 2\alpha = \frac{2r\sigma_x\sigma_y}{\sigma_x{}^2-\sigma_y{}^2}$$

where r = corr.(X,Y), $\sigma_x{}^2$ = Var(X) and $\sigma^2{}_y$ =Var(Y). Are U andV then independent?

(Do it your self)

Example:

Calculate Karl Pearson, coefficient of correlation from the following data

X	10	15	18	22	26	30	32	23
Y	8	10	12	14	16	20	16	24

Solution:

Calculation of correlation coefficient.

X	$(x-\bar{x})$	$(x-\bar{x})^2$	Y	$(y-\bar{y})$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
10	-12	144	8	-7	49	84
15	-7	49	10	-5	25	35
18	-4	16	12	-3	9	12

22	0	0	14	-1	1	0
26	4	16	16	1	1	4
30	8	64	20	5	25	40
32	10	100	16	1	1	10
23	1	1	24	9	81	9
$\sum x$=176		$\sum(x-\bar{x})^2$ =390	$\sum y = 120$		$\sum(y-\bar{y})^2$ =192	=194

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \times \sum(y-\bar{y})^2}}$$

$$\overline{X} = \frac{176}{8} = 22 \ : \ =22 \ : \ \overline{Y} = \frac{120}{8} = 15$$

$$r = \frac{194}{\sqrt{390 \times 192}} = \frac{194}{273.5} = .709$$

Hence we conclude that the correlation between X and Y is moderately high degree.

11.5 PROPERTIES OF CORRELATION COEFFICIENT:

1) Correlation coefficient is a pure number i.e. it has no limit
2) The correlation coefficient (or r) ranges from -1 to + 1
3) The correlation between two variables is known as simple correlation or correlation of zero order
4) It is not affected by coding (linear transformation) of two variables or variates values
5) The relation between the correlation coefficient r and the two regression correlation b_{xy} and b_{yx} is r = $\sqrt{b_{xy} b_{yx}}$
6) The sign of r will be same as that of b_{xy} and b_{yx}.
7) If the two variables are independent, the correlation coefficient between them is zero but the converse is not true.
8) If r =0 , it shows that the relationship between the two variables X & Y is not linear.

11.6 SELF ASSESSMENT QUESTION:

1) A computer while calculating the correlation coefficient between two variables X & Y obtained the following constants.

N= 25 , X =125, Y =100 , X^2=650 , Y^2=460 , XY=508

It was however later discovered at the time of checking that he had copied down two pairs of observations as

X:	6	8
Y:	14	6

While the correction values were

X:	8	6
Y:	12	8

Obtain the correct value of correlation coefficient between X & Y.

2) X_1 and X_2 are independent variables with mean 5 and 10 and standard deviations 2 and 3 respectively. Obtain r(U,V),

Where U= $3X_1 + 4X_2$ and V=$3X_1 - X_2$

3) If U = a X+ b Y & V = c X+ d Y , where X & Y are measured from their respective means and r is the correlation coefficient between X& Y , and if U & V are uncorrelated , show that

4) The following table gives the marks of Statistics and Economics of ten students. Calculate product moment correlation coefficient between the marks of mathematics and statistics and also interpret your results.

Marks of Statistics:	65	62	59	74	72	68	54	56	47	51
Marks of economics:	62	65	55	61	60	69	56	58	55	7

5.) Using formula : $\sigma_{x-y^2} = \sigma_{x^2} + \sigma_{y^2} - 2r(X,Y)\sigma_x\sigma_y$

Obtain the correlarion coefficient between the variables X and Y from the Following given data;

X:	65	66	67	68	69	70	71	67
Y:	67	68	64	72	70	67	70	68

****

UNIT - III **LESSON - 12**

SPEARMAN'S RANK CORRELATION COEFFICIENT

STRUCTURE

12.1 INTRODUCTION:

In the previous lesson we have studied about the correlation coefficient between two variables for which the observations are definitely measured.

When we are dealing with qualitative characteristics such as honesty, morality etc., which cannot be measured quantitatively but can be arranged serially. In such situations Karl Pearson's coefficient of correlation cannot be used.

Suppose we want to find if two characteristics A, say, intelligence and B, say beauty are related or not. Both the characteristics are incapable of quantitative measurements but we can arrange a group of n individuals in order of merit (ranks) w.r.t proficiency in the two characteristics. Let the random variables X and Y denotes the ranks of the individuals in the characteristics A and B respectively. If we assume that there is no tie, i.e. if no two individuals get the same rank in a characteristic then, obviously, X and Y assume numerical values ranging from 1 to n.

In another example, let us suppose that a group of n individuals is arranged in the order of merit or proficiency in two subjects 'Statistics and Mathematics'. Without any attempt to assess numerically assigning to each student a number which indicates his position in that group is known rank. In this way all the students are assigned with different ranks or numbers which indicates their positions in that group.

12.2 OBJECTIVES:

The main objectives of this lesson are:

1) To define Rank correlation coefficient
2) To calculate the Rank correlation coefficient when ranks are given
3) To calculate the Ranks correlation coefficient when ranks are not given
4) To calculate the Rank correlation coefficient when ranks are tied

12.3 SPEARMAN'S RANK CORRELATION COEFFICIENT:

This method was propounded by Prof. Spearman, hence it is known by his name. Calculation of correlation in this method is not done on the basis of actual data, but according to their place values (ranks).

The Pearsonian correlation coefficient between the ranks is called the rank correlation coefficient, usually denoted by ρ (rho) is given by the formula

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$\sum d^2$ = sum of squares of deviations between the rank of series.

n = number of items.

THEOREM: Prove that Spearman's formula for the rank correlation coefficient is given by

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

PROOF: Let (p_i, q_i), $(i = 1,2,........n)$ be the ranks of n individuals in the two characteristics A and B respectively.

Let $d_i = p_i - q_i$, denote the difference between the ranks of the i[th] individuals in the two characteristics.

If there is no tie (by assumption), then variables p and q takes values 1, 2 ....n.

$$\therefore \ \overline{p} = \overline{q} = \frac{n+1}{2} \qquad \text{and} \qquad \because \sigma_p^2 = \sigma_q^2 = \frac{n^2-1}{12}$$

We have $d_i = p_i - q_i = \left[(p-\overline{p})-(q-\overline{q})\right] \quad \because \ \overline{p} = \overline{q}$

$$\Rightarrow d_i^2 = (p_i-\overline{p})^2 + (q_i-\overline{q})^2 - 2(p_i-\overline{p})-(q_i-\overline{q})$$

Summing both sides over n values and dividing by n, we get

$$\frac{\sum d_i^2}{n} = \sigma_p^2 + \sigma_q^2 - 2\frac{\sum(p_i-\overline{p})-(q_i-\overline{q})}{n} \qquad (1)$$

$$\because \ \rho = \frac{\sum(p_i-\overline{p})-(q_i-\overline{q})}{n\sigma_p\sigma_q}$$

$$\Rightarrow \ \rho\,\sigma_p\sigma_q = \frac{\sum(p_i-\overline{p})-(q_i-\overline{q})}{n} \qquad (2)$$

Substituting (2) in (1), we get

$$\frac{\sum d_i^2}{n} = \sigma_p^2 + \sigma_q^2 - 2\,\rho\sigma_p\sigma_q$$

$$= \ 2\,\sigma_p^2(1-\rho) \qquad \because \sigma_p = \sigma_q$$

$$\Rightarrow (1-\rho) = \frac{\sum d_i^2}{2\,n\sigma_p^2} = \frac{6\sum d_i^2}{n(n^2-1)}$$

$$\Rightarrow \ \rho = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

THEOREM: Show that Rank Correlation Coefficient lies between -1 to +1.

PROOF: we know that

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} \qquad \text{(A)}$$

If $p_i = q_i$, then $d_i = 0$

hence $\rho = 1$ (1)

i.e. ρ will be maximum if $\sum d_i^2$ is minimum.

Again ρ will be minimum if $\sum d_i^2$ is maximum and the maximum value of

$\sum d_i^2$ is $\frac{n(n^2-1)}{3}$, then

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} \qquad \text{(2)}$$

Substitute the maximum value of $\sum d_i^2$ in (A) we get

$$\rho = 1 - \frac{6\left[\frac{n(n^2-1)}{3}\right]}{n(n^2-1)}$$

$= -1$

Hence minimum value of $\rho = -1$ (3)

From (1) and (3), we get

$-1 \le \rho \le 1$.

<u>When ranks are Tied:</u>

If some of the individuals receive the same rank, they are said to be tied, then Spearman's formula for calculating the rank correlation coefficient breaks down.

In this case, common ranks are assigned to the repeated items. These common ranks are the arithmetic mean of the ranks.

For example, suppose an item is repeated at rank 6. Then the common rank to be assigned to each item is $\frac{6+7}{2}$ i.e. 6.5, which is the average of 6 and 7. The next item will be assigned the rank 8.

The formula for correction of Rank Correlation is thus modified in the following manner

$$\rho = 1 - \frac{6\left[\sum d^2 + \frac{1}{12}\left(m^3 - m\right) + \frac{1}{12}\left(m^3 - m\right) + ...\right]}{n^2\left(n^2 - 1\right)}$$

Where m is the number of times an item is repeated.

12.4 MERITS OF RANK CORRELATION COEFFICIENT :

1) Since in this method or the sum of the difference be R and Ris always equal to zero, it provides a check on the calculations.

2) Since Spearman's Rank Correlation Coefficient is the same thing ascending Karl Pearson's Coefficient of correlation between ranks. It can be interpreted in the same way as Karl Pearson's coefficient of correlation .

3) Rank Correlation unlike Karl Pearson's coefficient of correlation does not assume normality in the universe from which the sample has been taken.

4) Rank correlation is very easy to understand and apply.

5) Spearman's Rank method is the only way of studying correlation between qualitative data which cannot measure in figures but can be arranged in serial order.

12.5 DEMERITS OF RANK CORRELATION COEFFICIENT:

1. The method cannot be used in two-way frequency table or bivariate frequency distribution.

   It can be conveniently used only when n is small say 30, otherwise calculations become tedious.

EXAMPLE 1: Find coefficient of correlation by ranking method:

m:	12	10	14	16	18	20	22	25	26	23
n:	8	13	4	15	6	3	9	10	1	7

SOLUTION:

m	n	Rank in m	Rank in n	d= m-n	d^2
12	8	9	5	4	16
10	13	10	2	8	64
14	4	8	8	0	0
16	15	7	1	6	36
18	6	6	7	-1	1
20	3	5	9	-4	16
22	9	4	4	0	0
25	10	2	3	-1	1
26	1	1	10	-9	81
23	7	3	6	-3	9

n=10 $\sum d^2$ =224

$$\rho = 1-\frac{6\sum d^2}{n(n^2-1)}$$

$$= 1-\frac{6\times 224}{10(10^2-1)}$$

$$= 1-\frac{1344}{10(100-1)}$$

$$=1-\frac{1344}{10(99)}$$

$$=1-\frac{1344}{990}$$

$$=1-1.36$$

$$=-0.36$$

EXAMPLE 2: Find coefficient of correlation by Spearman's ranking method:

m:	13	10	15	16	14	15	10	8
n:	6	8	10	8	4	5	6	3

SOLUTION:

m	n	Rank in m	Rank in n	d=m-n	d^2
13	6	5	4.5	0.5	0.25
10	8	6.5	2.5	4	16.00
15	10	2.5	1.0	1.5	2.25
16	8	1	2.5	-1.5	2.25
14	4	4	7.0	-3	9.00
15	5	2.5	6.0	-3.5	12.25
10	6	6.5	4.5	2	4.00
8	3	8	8.0	0	0.00

n=8 $\sum d^2 = 46.00$

In m-series, the value 15 occurs two time and its rank is 2.5 i.e., the average of 2 and 3.

Again in m-series, the value 10 occurs two time and its rank is 6.5 i.e., the average of 6 and 7

In n-series, the value 8 occurs two time and its rank is 2.5 i.e., the average of 2 and 3

Again in n-series, the value 6 occurs two time and its rank is 4.5 i.e., the average of 4 and 5.

$$\therefore \quad \rho = 1 - \frac{6\left[\sum d^2 + \frac{1}{12}\left(m^3 - m\right) + \frac{1}{12}\left(m^3 - m\right) + \ldots\ldots\right]}{n^2\left(n^2 - 1\right)}$$

$$= 1 - \frac{6\left[46 + \frac{1}{12}\left(2^3 - 2\right) + \frac{1}{12}\left(2^3 - 2\right) + \frac{1}{12}\left(2^3 - 2\right) + \frac{1}{12}\left(2^3 - 2\right)\right]}{8^2\left(8^2 - 1\right)}$$

$$= \quad 1 - \frac{6 \times 48}{504}$$

$$= \quad 1 - 0.57 \qquad = 0.43$$

12.6 SELF ASSESSMENT:

1) The following data provides the ranks of 10 students in Statistics and Economics

Statistics:	4	5	7	8	10	1	3	6	2	9
Economics:	3	4	7	9	10	8	6	5	2	1

Calculate the rank correlation co-efficient. (Ans: 0.24)

2) Calculate the coefficient of correlation for ranks from the following data:

(X,Y) : (78,84), (36,51), (98,91)(25,60), (75,68), (82,62), (90,86), (62,58), (65,63), (39,47)

Ans: 0.8

3) Ten competitors in a music test were ranked by the three judges x,y and z in the following order:

Ranks by x:	1	6	5	10	3	2	4	9	7	8

Ranks by y:	3	5	8	4	7	10	2	1	6	9
Ranks by z:	6	4	9	8	1	2	3	10	5	7

Using rank correlation method, discuss which pair of judges has the nearest approach to common likings in music.

4) A sample of 12 fathers and their eldest sons gave the following data about their height in inches:

Father:	65	63	67	64	68	62	70	66	68	67	69	71
Son:	68	66	68	65	69	66	68	65	67	68	70	

Calculate coefficient of rank correlation (Ans. 0.72)

5) The ranking of 10 persons at the end of training are as follows:

Rank before:	3	8	5	11	7	4	9	12	10	6
Rank after :	8	10	5	9	4	3	7	8	6	12

Calculate spearman's coefficient of correlation.

****

UNIT-III **LESSON -13**

CORRELATION RATIO

STRUCTURE

13.1 INTRODUCTION:

As discussed earlier, when variables are linearly related, we have regression lines of one variable on another variable and correlation coefficient can be computed to tell us about the extent of association between them. However, if the variables are not linearly related but some sort of curvilinear relationship exists between them, the use of r which is a measure of the degree to which the relation approaches straight line "law" will be misleading.

We might come across bivariate distributions where 'r' may be very low or even zero but the regression may be strong, or even perfect. Correlation ratio η is the appropriate measure of curvilinear relationship between the two variables.

13.2 MEASURE OF CORRELATION RATIO:

We have assumed that there is a single observed value Y corresponding to the given value x_i of X but sometimes there are more than one such value of Y.

Suppose corresponding to the value x_i, $i = 1,2,.....m$ of the variable X, the variable Y takes the values y_{ij} with respective frequencies f_{ij}; $j = 1,2,....n$.

Though all the $x's$ in the ith vertical array have the same value, the $y's$ are different. A typical pair of values in the ith array is (x_i, y_{ij}), with frequency f_{ij}. Thus the first suffix I indicates the vertical array while the second suffix j indicates the positios of y in that array. Let

$$\sum_{j=1}^{n} f_{ij} = n_i \quad \text{and} \quad \sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} f_{ij}\right) = \sum_{i=1}^{m} n_i = N \qquad \text{(say)}$$

if $\bar{y}_i$ and $\bar{y}$ denote the means of the ith array and the overall mean respectively, then

$$\bar{y}_i = \frac{\sum_{j=1}^{n} f_{ij}\, y_{ij}}{\sum_{j=1}^{n} f_{ij}} = \frac{\sum_{j=1}^{n} f_{ij}\, y_{ij}}{n_i} = \frac{T_i}{n_i} \text{ and}$$

$$\bar{y} = \frac{\sum_i \sum_j f_{ij}\, y_{ij}}{\sum_i \sum_j f_{ij}} = \frac{\sum_i n_i \bar{y}_i}{\sum_i n_i} = \frac{T}{N}$$

In other words $\bar{y}$ is the weighted mean of all the array means, the weights being the array frequencies.

13.3 DEFINITION:

The Correlation ration of Y on X, usually denoted by η_{YX} is given by:

$$\eta_{YX}{}^2 = 1 - \frac{\sigma_{eY}{}^2}{\sigma_Y^2},$$

where , $\sigma_{eY}^2 = \frac{1}{N}\sum_i \sum_j f_{ij}\left(y_{ij} - \bar{y}_i\right)^2$ and $\sigma_Y^2 = \frac{1}{N}\sum_i \sum_j f_{ij}\left(y_{ij} - \bar{y}\right)^2$

A convenient expression for η_{YX} can be obtained in terms of standard deviation σ_{mY} of the means of vertical arrays, eavh mean being weighted by the array frequency. We have

$$N\sigma_Y^2 = \sum_i \sum_j f_{ij}\left(y_{ij} - \bar{y}\right)^2$$

$$= \sum_i \sum_j f_{ij}\left\{\left(y_{ij} - \bar{y}_i\right) + \left(\bar{y}_i - \bar{y}\right)\right\}^2$$

$$\therefore \quad \eta_{YX}{}^2 = \left[\sum_i \left(\frac{T_i^2}{n_i}\right) - \frac{T^2}{N}\right] \Big/ N\sigma_Y^2$$

a formula, much more convenient for computational purposes.

13.4 INTRA–CLASS CORRELATION:

Intra–class correlation means within class correlation. It is distinguishable from product moment correlation in as much as here both the variables measure the same characteristics.

Sometimes specially in biological and agricultural study, it is of interest to know how the members of a family or group are correlated among themselves with respect to some one of their common characteristics.

For example, we may require the correlation between the heights of brothers of a family or between yields of plots of an experimental block. In such cases both the variables measure the same characteristics, e.g., height and height or weight and weight. There os nothing to distinguish one from the others so that one may be treated as X–variable and the other as the Y–variable.

Suppose we have $A_{1,} A_2 \ldots$An families with $k_{1,} k_2 \ldots k_n$ members, each of which may be represented as shown below:

$$\begin{array}{llll} x_{11} & x_{21} \;\ldots\ldots\ldots & x_{i1} \ldots\ldots\ldots & x_{n1} \\ x_{12} & x_{22} \ldots\ldots\ldots & x_{i2} \ldots\ldots\ldots & x_{n2} \\ . & . & . & . \end{array}$$

$$
\begin{array}{llll}
. & . & . & . \\
. & . & . & . \\
x_{1j} & x_{2j}\ldots\ldots\ldots & x_{ij}\ldots\ldots\ldots & x_{nj} \\
x_{1k_1} & x_{2k_2}\ldots\ldots\ldots & x_{ik_j}\ldots\ldots\ldots & x_{nk_n}
\end{array}
$$

and let x_{ij} $(i=1,2,......,n;\ j=1,2,.....k_i)$ denote the measure ment on the jth member in the ith family.

We shall have $k_i\,(k_i-1)$ pairs for the ith family or group like $(x_{ij}, x_{il}),\ j \neq l$. There will be $\sum k_i\,(k_i-1)$ = N, pairs for all the n families or groups. If we prepare a correlation table there will be $k_i\,(k_i-1)$ entries for the ith group or family and $\sum k_i\,(k_i-1)$ = N entries for all the n families or groups.

The table is symmetrical about the principal diagonal. Such a table is called an intra-class correlation table and the correlation is called intra-class correlation.

Since $$\bar{x} = \bar{y} = \frac{1}{N}\left[\sum_i\left\{(k_i-1)\sum_j x_{ij}\right\}\right]$$

Similarly, $$\sigma_x^2 = \sigma_y^2 = \frac{1}{N}\left[\sum_i\left\{(k_i-1)\sum_j (x_{ij}-\bar{x})^2\right\}\right]$$

Further $$\text{Cov(X,Y)} = \frac{1}{N}\sum_i\left[\sum_{j,l}(x_{ij}-\bar{x})(x_{il}-\bar{x})\right],\ j \neq l$$

$$= \frac{1}{N}\sum_i\left[\sum_{j=1}^{k_i}\sum_{l=1}^{k_i}(x_{il}-\bar{x})(x_{ij}-\bar{x}) - \sum_{j=1}^{k_i}(x_{ij}-\bar{x})^2\right]$$

If we write $\bar{x}_i = \sum_j x_{ij}/k_i$, then

$$\sum_i\left[\sum_{j=1}^{k_i}\sum_{l=1}^{k_i}(x_{il}-\bar{x})(x_{ij}-\bar{x}) - \sum_{j=1}^{k_i}(x_{ij}-\bar{x})^2\right] = \sum_i k_i^2\,(\bar{x}_i-\bar{x})^2$$

Therefore intra-class correlation is given by

$$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$$

$$= \frac{\sum_i k_i^2 (\bar{x}_i - \bar{x})^2 - \sum_i \sum_j (x_{il} - \bar{x})^2}{\sum_i \sum_j k_i (k_i - 1)(x_{ij} - \bar{x})^2}$$

If we put $k_i = k$, then

$$r = \frac{1}{(k-1)} \left\{ \frac{k\sigma_m^{\ 2}}{\sigma^2} - 1 \right\}$$

Where σ^2 denotes the variance of X and σ_m^2 the variance of means of families.

13.5 LIMITS OF INTRA-CLASS CORRELATION:

$$-\frac{1}{(k-1)} \leq r \leq 1.$$

INTERPRETATION:

Intra-class correlation cannot be less than $-\frac{1}{(k-1)}$, though it may attain the value +1 on the positive side, so that it is a skew coefficient and a negative value has not the same significance as a departure from independence as an equivalent positive value.

Example: In five families of three each the heights in inches of the brothers are (69,70,71),(70,71,72),(71,72,73),(72,73,74),(73.74,75).

Find the intra class correlation coefficient.

Solution: If the number of members in each family is k, then

$$r = \frac{1}{(k-1)} \left\{ \frac{k\sigma_m^{\ 2}}{\sigma^2} - 1 \right\}$$

The sample means $\bar{x}_i$ if five families are (70,71,72,73,74).

Hence the mean of these means ($\bar{x}$) is 72.

Variance (σ_m^2) of the means of the families is 2.

Common variance (σ^2) = $\frac{8}{3}$

$\therefore$ r = 0.625.

13.6 COEFFICIENT OF DETERMINATION:

Coefficient of correlation between two variable series is a measure of linear relationship between them and indicates the amount of variation of one variable which is associated with or is accounted for by another variable.

A more useful and readily comprehensible measure for this purpose is the coefficient of determination which gives the percentage variation in the dependent variable that is accounted for by the independent variable.

In other words, the coefficient of determination gives the ratio of the explained variance to the total variance. The coefficient of determination is given by the square of the correlation coefficient, i.e., r^2 . Thus,

$$\text{Coefficient of determination} = r^2 = \frac{Explained\ Variance}{Total\ Variance}.$$

The coefficient of determination is a much useful and better measure for interpreting the value of r.

Coefficient of determination is always non-negative.

<u>Coefficient of Non-determination</u>: The ratio of the unexplained variation to the total variation is called the coefficient of non-determination. It is usually denoted by K^2 and is given by the formula:

$$K^2 = \frac{Un-\exp lained\ Variance}{Total\ Variance} = 1 - r^2.$$

13.7 SELF ASSESSMENTS:

1) Define correlation ratio.

2) When the correlation coefficients is equal to unity, show that the two correlation ratios are also equal to unity.

3) What do you understand by intra-class correlation coefficient?

4) What is the coefficient of determination? How is it useful in interpreting the value of an observed correlation coefficient ?

****

UNIT-IV **LESSON-14**

RANDOM EXPERIMENT

14.1. Introduction

Many kinds of investigations may be characterized in part by the fact that repeated experimentation, under essentially the same conditions, is more or less standard procedure. For instance, in medical research, interest may center on the effect of a drug that is to be administered; or an economist may be concerned with the prices of three specified commodities at various time intervals; or the agronomist may wish to study the effect that a chemical fertilizer has on the yield of a cereal grain. The only way in which an investigator can elicit information about any such phenomenon is to perform his experiment. Each experiment terminates with an outcome. But it is characteristic of these experiments that the outcome cannot be predicted with certainty prior to the performance of the experiment. Probability Theory involves study of random experiments or phenomenon.In this lesson we define the terms and definations associated with the concept of Probability.Suitable examples have been quoted to explain the terms and make things simpler.Set theoretic concepts of events have been introduces in the lesson.

14.2. Objectives

1. To introduce the concept of random experiment
2. To introduce the definations of sample space and events
3. To introduce the basic concepts related to Probability Theory.

14.3. Sample spaces

The world around us is full of phenomena we perceive as random or unpredictable. We model these phenomena as outcomes of some experiment, where you should think of experiment in a very general sense. The outcomes are elements of a sample space Ω, and subsets of Ω are called **events.** The events will be assigned a probability, a number between 0 and 1 that expresses how likely the event is to occur.

Suppose that we have such an experiment, the outcome of which cannot be predicted with certainty, but the experiment is of such a nature that the collection of every possible outcome can be described prior to its performance. If this kind of experiment can be repeated under the same conditions, it is called a **random experiment**, and the collection of every possible outcome is called the experimental space or the **sample space**.

***Random Experiment*:** An experiment is a random experiment if its outcome cannot be predicted precisely. One out of a number of outcomes is possible in a random experiment. A single performance of the random experiment is called a *trial*.

Sample spaces are simply sets whose elements describe the all possible outcomes of the experiment in which we are interested.

We start with the most basic experiment: the tossing of a coin. Assuming that we will never see the coin land on its rim, there are two possible outcomes: heads and tails. We therefore take as the sample space associated with this experiment the set Ω = {H, T}.

In another experiment we ask the some person in which month his birthday falls. An obvious choice for the sample space is all month names in a year.

Ω = (Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec)

In a third experiment we put load on a bridge up to the point where the structure collapses. The outcome of this experiment is the load at which the structure collapses. In reality, one can only measure with finite accuracy, say up to five decimals, and a sample space with just those numbers would strictly be adequate. However, in principle, the load itself could be any positive number and therefore Ω = (0, 8) is the correct choice. Even though in reality there may also be an upper limit to what loads are conceivable, it is not necessary or practical to try to limit the outcomes correspondingly.

In a fourth experiment, we find on our doormat three envelopes, sent to us by three different persons, and we look in which order the envelopes lie on top of each other. Coding them 1, 2, and 3, the sample space would be

$$\Omega = \{123, 132, 213, 231, 312, 321\}.$$

Example 14.3.1 If we received mail from four different persons, how many elements would the corresponding sample space have ?

In general one might consider the order in which n different objects can be placed. This is called a permutation of the n objects. As we have seen, there are 6 possible permutations of 3 objects, and $4 \cdot 6 = 24$ of 4 objects.

What happens is that if we add the nth object, then this can be placed in any of n positions in any of the permutations of n - 1 objects. Therefore there are

$$n \cdot (n-1) \cdot \cdots \cdot 3 \cdot 2 \cdot 1 = n!$$

possible permutations of n objects. Here n! is the standard notation for this product and is pronounced "n factorial." It is convenient to define 0! = 1.

14.4 Events and Algebra of Events

Subsets of the sample space are called events . We say that an event 'A' occurs if the outcome of the experiment is an element of the set A. For example, in the birthday experiment we can ask for the outcomes that correspond to a long month, i.e., a month with 31 days. This is the event

$$L = \{Jan, Mar, May, Jul, Aug, Oct, Dec\}.$$

Events may be combined according to the usual set operations.

For example if R is the event that corresponds to the months that have the letter r in their (full) name (so R = {Jan, Feb, Mar, Apr, Sep, Oct, Nov, Dec}), then the long months that contain the letter r are

$L \cap R = \{Jan, Mar, Oct, Dec\}.$

The set $L \cap R$ is called the intersection of L and R and occurs if both L and R occur.

Similarly, we have the union $A \cup B$ of two sets A and B , which occurs if at least one of the events A and B occurs. Another common operation is taking complements. The event $A^c = \{\omega \in \Omega : \omega \in A\}$ is called the complement of A; it occurs if and only if

A does not occur.

The complement of ω is denoted Ø, the empty set, which represents the impossible event.

We call events A and B disjoint or mutually exclusive if A and B have no outcomes in common; in set terminology: $A \cap B = \emptyset$. For example, the event L "the birthday falls in a long month" and the event {Feb} are disjoint.

Finally, we say that event A implies event B if the outcomes of A also lie in B . In set notation: A ? B ; Some people like to use double negations:

"It is certainly not true that neither John nor Mary is to blame."

This is equivalent to: "John or Mary is to blame, or both." The following useful rules formalize this mental operation to a manipulation with events.

DeMorgan's laws. For any two events A and B we have

$(A \cap B)^c = A^c \cup B^c$ and $(A \cup B)^c = A^c \cap B^c$.

Exercise 1.4.1 Let J be the event "Ram speaks the truth" and M the event "Rahim speaks the truth." Express the two statements above in terms of the events J, J^c, M, and M^c, and check the equivalence of the statements by means of DeMorgan's laws.

Further we can explore the concept of events in general.An event is defined as any outcome that can occur. There are two main categories of events: Deterministic and Probabilistic. A deterministic event always has the same outcome and is predictable 100% of the time.

- Distance traveled = time x velocity
- The speed of light
- The sun rising in the east
- James Bond winning the fight without a scratch

A probabilistic event is an event for which the exact outcome is not predictable 100% of the time.

- The number of heads in ten tosses of a coin

- The winner of the World Series
- The number of games played in a World Series
- The number of defects in a batch of product

In a boxing match there may be three possible events. (There could be more depending on the question asked.)

- Fighter A wins
- Fighter B wins
- Draw

Four Basic Types of Events

- ***Mutually Exclusive Events:*** These are events that cannot occur at the same time. The cause of mutually exclusive events could be a force of nature or a man made law. Being twenty-five years old and also becoming president of the United States are mutually exclusive events because by law these two events cannot occur at the same time.

- ***Complementary Events:*** These are events that have two possible outcomes. The probability of event A plus the probability of A equals one. P(A) + P(A') = 1. Any event A and its complementary event A' are mutually exclusive. Heads or tails in one toss of a coin are complementary events.

- ***Independent Events:*** These are two or more events for which the outcome of one does not affect the other. They are events that are not dependent on what occurred previously. Each toss of a fair coin is an independent event.

Conditional Events: These are events that are dependent on what occurred previously. If five cards are drawn from a deck of fifty-two cards, the likelihood of the fifth card being an ace is dependent on the outcome of the first four cards.

14.5 Probability

We want to express how likely it is that an event occurs. To do this we will assign a probability to each event. The assignment of probabilities to events is in general not an easy task, and some of the following lessons will be dedicated directly or indirectly to this task. Since each event has to be assigned a probability, we speak of a probability function.

The probability of an event A is a number $P(A)$ assigned to the event . Let us see how we can define probability.

14.5.1.Classical definition of probability (Laplace 1812)

Consider a random experiment with a finite number of outcomes N. If all the outcomes of the experiment are *equally likely*, the probability of an event A is defined by

$$P(A)=\frac{N_A}{N}$$

Where N_A = Number of outcomes favourable to *A*.

Example 1.5.1 A fair die is rolled once. What is the probability of getting a'6'?

Here $S=\{'1',\ '2',\ '3',\ '4',\ '5',\ '6'\}$ and $A=\{\ '6'\}$

$$\therefore N=6 \text{ and } N_A=1$$

$$\therefore P(A)=\frac{1}{6}$$

Example 14.5.2 A fair coin is tossed twice. What is the probability of getting two 'heads'?

Here $S=\{HH,\ TH,\ TT,\ TT\}$ and $A=\{HH\}$.

Total number of outcomes is 4 and all four outcomes are equally likely.

Only outcome favourable to A is {HH}

$$\therefore P(A)=\frac{1}{4}$$

Remarks

- The classical definition is limited to a random experiment which has only a finite number of outcomes. In many experiments like that in the above example, the sample space is finite and each outcome may be assumed 'equally likely.' In such cases, the *counting method* can be used to compute probabilities of events.

- Consider the experiment of tossing a fair coin until a 'head' appears.As we have discussed earlier, there are countably infinite outcomes. Can you believe that all these outcomes are equally likely?
- The notion of equally likely is important here. Equally likely means equally.

probable. Thus this definition presupposes that all events occur with equal *probability*. Thus the definition includes a concept to be defined.

14.5.2. Relative-frequency based definition of proability(von Mises, 1919)

If an experiment is repeated n times under similar conditions and the event A occurs in n_A times, then

$$P(A) = \lim_{n\to\infty} \frac{n_A}{n}$$

Example 14.5.3. Suppose a die is rolled 500 times. The following table shows the frequency each face.

Face	**1**	**2**	**3**	**4**	**5**	**6**
Frequency	**82**	**81**	**88**	**81**	**90**	**78**
Relative frequency	**0.164**	**0.162**	**0.176**	**0.162**	**0.18**	**0.156**

In the above table, we see that relative frequencies are close to $\frac{1}{6}$. How do we ascertain that these relative frequencies will approach $\frac{1}{6}$ as a limit as $n \to \infty$?

Remarks

The relative-frequency based definition is also inadequate from the theoretical point of view.

- We cannot repeat an experiment infinite number of times.
- We cannot ascertain that the above ratio will converge for all possible sequences of outcomes of the experiment.

14.5.3. Axiomatic definition of probability (Kolmogorov, 1933)

We have earlier defined an event as a subset of the sample space. *Does each subset of the sample space forms an event?*

The answer is *yes* for a finite sample space. However, we may not be able to assign probability meaningfully to all the subsets of a continuous sample space. We have to eliminate those subsets. The concept of the *sigma algebra* is meaningful now.

Definition: Let S be a sample space and F a sigma field defined over it. Let $P : \mathrm{F} \to \mathbb{R}$ be a mapping from the sigma-algebra F into the real line such that for each $A \in \mathrm{F}$, there exists a unique $P(A) \in \mathbb{R}$. Clearly P is a set function and is called probability i1 A t satisfies the following axioms

A

1. $P(A) \geq 0$ for all $A \in \mathrm{F}$
2. $P(S) = 1$
3. Countable additivity If $A_1, A_2, \ldots$ are pair-wise disjoint events, i.e. $A_i \cap A_j = \phi$ for $i \neq j$, then

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

In other words,we can define the Probability of an event as also

Definition. A probability function P on a finite sample space *S* assigns to each event A in *S* ,a number P(A) in [0,1] such that

(i) P(*S*) = 1, and

(ii) P(A $\cup$ B) = P(A) + P(B) if A and B are disjoint.

The number P(A) is called the probability that event A occurs.

Property (i) expresses that the outcome of the experiment is always an element of the sample space, and property (ii) is the additivity property of a probability function. It implies additivity of the probability function over more than two sets; e.g., if A, B , and C are disjoint events, then the two events A ? B and C are also disjoint, so

P(A $\cup$ B $\cup$ C) = P(A $\cup$ B) + P(C) = P(A) + P(B) + P(C) .

We will now look at some examples. When we want to decide whether Peter or Paul has to wash the dishes, we might toss a coin. The fact that we consider this a fair way to decide translates into the opinion that heads and tails are equally likely to occur as the outcome of the coin-tossing experiment. So we put

$$P(\{H\}) = P(\{T\}) = 1/2$$

Formally we have to write {H } for the set consisting of the single element H, because a probability function is defined on events, not on outcomes. From now on we shall drop these brackets.

Now it might happen, for example due to an asymmetric distribution of the mass over the coin, that the coin is not completely fair. For example, it might be the case that

$P(H) = 0.4999$ and $P(T) = 0.5001$.

More generally we can consider experiments with two possible outcomes, say "failure" and "success", which have probabilities $1 - p$ and p to occur, where p is a number between 0 and 1. For example, when our experiment consists of buying a ticket in a lottery with 10 000 tickets and only one prize, where "success" stands for winning the prize, then $p = 10^{-4}$.

How should we assign probabilities in the second experiment, where we ask for the month in which the next person we meet has his birthday? In analogy with what we have just done, we put

$$P(Jan) = P(Feb) = \cdots = P(Dec) = 1/12.$$

Some of you might object to this and propose that we put, for example,

$$P(Jan) = 31/365 \quad \text{and} \quad P(Apr) = 30/365 ,$$

because we have long months and short months. But then the very precise among us might remark that this does not yet take care of leap years.

Exercise 14.5.4. If you would take care of the leap years, assuming that one in every four years is a leap year (which again is an approximation to reality!), how would you assign a probability to each month?

In the third experiment (the putting of load on a bridge), where the outcomes are real numbers, it is impossible to assign a positive probability to each outcome (there are just too many outcomes!). But we shall be restricting ourselves here to finite and countably infinite sample spaces.

In the fourth experiment it makes sense to assign equal probabilities to all six outcomes :

$P(123) = P(132) = P(213) = P(231) = P(312) = P(321) = 1/6$

Until now we have only assigned probabilities to the individual outcomes of the experiments. To assign probabilities to events we use the additivity property. For instance, to find the probability $P(T)$ of the event T that in the three envelopes experiment envelope 2 is on top we note that

$$P(T) = P(213) + P(231) = 1/6 + 1/6 = 1/3 .$$

In general, additivity of P implies that the probability of an event is obtained by summing the probabilities of the outcomes belonging to the event.

Excercise 14.5.5 . Compute P(L) and P(R) in the birthday experiment.

Finally we mention a rule that permits us to compute probabilities of events A and B that are not disjoint. Note that we can write

$$A = (A \cap B) \cup (A \cap B^c),$$

which is a disjoint union; hence

$P(A) = P(A \cap B) + P(A \cap B^c)$.

If we split $A \cap B$ in the same way with B and B^c , we obtain the events $(A \cup B) \cap B$, which is simply B and $(A \cup B) \cap B^c$, which is nothing but $A \cap B^c$.

1 This means: although infinite, we can still count them one by one; $\Omega = \{\omega 1 , \omega 2 , . . . \}$. The interval [0,1] of real numbers is an example of an uncountable sample space.

Thus

$P(A \cup B) = P(B) + P(A \cap B^c)$.

Eliminating $P(A \cap B^c)$ from these two equations we obtain the following rule.

The probability of a union. For any two events A and B we have

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

From the additivity property we can also find a way to compute probabilities of complements of events: from $A \cup A^c = S$, we deduce that

$$P(A^c) = 1 - P(A) .$$

14.6 Products of sample spaces

Basic to statistics is that one usually does not consider one experiment, but that the same experiment is performed several times. For example, suppose we throw a coin two times. What is the sample space associated with this new experiment? It is clear that it should be the set

$$\Omega = \{H, T\} \times \{H, T\} = \{(H, H), (H, T), (T, H), (T, T)\}.$$

If in the original experiment we had a fair coin, i.e., $P(H) = P(T)$, then in this new experiment all 4 outcomes again have equal probabilities:

$$P((H, H)) = P((H, T)) = P((T, H)) = P((T, T)) = 1/4.$$

Somewhat more generally, if we consider two experiments with sample spaces Ω_1 and Ω_2 then the combined experiment has as its sample space the set

$$\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \omega\Omega_1, \omega_2 \in \omega_2\}.$$

If Ω_1 has r elements and Ω_2 has s elements, then $\Omega_1 \times \Omega_2$ has rs elements. Now suppose that in the first, the second, and the combined experiment all outcomes are equally likely to occur. Then the outcomes in the first experiment have probability 1/r to occur, those of the second experiment 1/s, and those of the combined experiment probability 1/rs. Motivated by the fact that $1/rs = (1/r) \times (1/s)$,

we will assign probability p_ip_j to the outcome (ω_i, ω_j) in the combined experiment, in the case that ω_i has probability p_i and ω_j has probability p_j to occur. One should realize that this is by no means the only way to assign probabilities to the outcomes of a combined experiment.

The preceding choice corresponds to the situation where the two experiments do not influence each other in any way. What we mean by this influence will be explained in more detail in the next lesson.

14.7. Lesson End Excercises

1. Let A and B be events. Find an expression and exhibit the Venn diagram for the event that (i) A or not B occurs, (ii) neither A nor B occurs.

2. Let A, B and C be events. Find an expression and exhibit the Venn diagram for the event that

   (i) exactly one of the three events occurs,(ii) at least two of the events occurs,(iii) none of the events occurs,(iv) A or B, but not C, occurs.

3. Let a penny, a dime and a die be tossed.

(i) Describe a suitable sample space S.

(ii) Express explicitly the following events :

A = (two heads and an even number appear),

B = (a 2 appears), C = (exactly one head and a prime number appear),

(iii) Express explicitly the event that (a) A and B occur, (b) only B occurs,(c) B or C occurs.

4. Three students A, B and C are in a swimming race. A and B have the same probability of winning and each is twice as likely to win as C. Find the probability that B or C wins.

5. A die is weighted so that the even numbers have the same chance of appearing, the odd numbers have the same chance of appearing, and each even number is twice as likely to appear as any odd number. Find the probability that (i) an even number appears, (ii) a prime number appears, (iii) an odd number appears, (iv) an odd prime number appears.

6. One card is selected at random from 50 cards numbered 1 to 60. Find the probability that the number on the card is (i) divisible by 5,(ii) prime,(iii) ends in the digit 2.

7. Of 10 girls in a class, 3 have blue eyes. If two of the girls are chosen at random,

what is the probability that (i) both have blue eyes,(ii) neither has blue eyes,(iii) at least one has blue eyes?

8. Three bolts and three nuts are put in a box. If two parts are chosen at random, find the probability that one is a bolt and one a nut.

1.8. Suggested Readings

1. Goon,Gupta;Das Gupta (1991).Fundamental of Statistics
2. Gupta,S.C. and Kapoor,V.K. Fundamental of Mathematical Statistics
3. Hoel,P.G. (1971).Introductory of Mathematical Statistics
4. Hogg R.V and Craig,A.T.Introduction to Mathematical Statistics
5. Hogg ,RV and Tanis,EA(1993).Probability and Statistical Inference
6. Mood,AM,Bose DC and Graybill F A.Introduction to the Theory of Statistics
7. Rohtagi,VK.An Introduction to Probability Theory and Mathematical Statistics

****

UNIT-IV **LESSON-15**

AXIOMS OF PROBABILITY

15.1 Introduction

In continuation of the last lesion ,we extend the definition of probability in new domain. In this lesson we introduce the most communally used definition of Probability in Statistics i.e. axiomatic definition of Probability.Also we have proved certain theorems useful in solving the problems and further understanding of advance concepts.

15.2. Objectives

1. To introduce the Axiomatic definition of Probablity
2. To establish the Standard theorems based on proabability
3. To illustrate the concepts with some examples.

15.3. AXIOMS OF PROBABILITY

Let S be a sample space, let E be the class of events, and let P be a real-valued function defined on Ω. Then P is called a probability function, and P(A) is called the probability of the event A if the following axioms hold:

[P1] For every event A, $0 = P(A) \leq 1$.

[P2] $P(S) = 1$.

[P3] If A and B are mutually exclusive events, then

$$P(A \cup B) = P(A) + P(B)$$

[P4] If A1, A2, ... is a sequence of mutually exclusive events, then

$$P(A1 \cup A2 \cup \cdots) = P(A1) + P(A2) + \ldots\ldots\ldots$$

The following remarks concerning the axioms [P3] and [P4] are in order. First of all, using (P_3) and mathematical induction we can prove that for any mutually exclusive events A1,A2, ...,An, We emphasize that [P4] does not follow from [P3] even though (*) holds for every positive integer n. However, if the sample space S is finite, then clearly the axiom [P4] is superfluous.

Remark

- The triplet (S, F, P). is called the *probability space*.
- Any assignment of probability assignment must satisfy the above three axioms
- If $\quad if\ A \cap B = \emptyset, \quad P(A \cup B) = P(A) + P(B)$

  This is a special case of axiom 3 and for a *discrete sample space*, this simpler version may be considered as the axiom 3. We shall give a proof of this result below.

  The events A and B are called *mutually exclusive*

  We now prove a number of theorems which follow directly from our axioms.

Theorem15.3.1: If Ø is the empty set, then P(Ø) = 0.

Proof: Let A be any set; then A and Ø are disjoint and (AUØ) = A.

By [P3],

P(A) = P(A U Ø) = P(A) + P(Ø)

Subtracting P(A) from both sides gives our result.

Theorem 2.3.2: If A^c is the complement of an event A, then $P(A^c) = 1 - P(A)$.

Proof: The sample space S can be decomposed into the mutually exclusive events A and A^c; that is,

$S = A \, UA^c$. By [P2] and [P3] we obtain

$1= P(S) = P(A \, UA^c) = P(A) + P(A^c)$

from which our result follows.

Theorem15.3.3: If A ?B, then P(A) =P(B).

Proof. If A ?B, then B can be decomposed into the mutually

exclusive events A and B \ A

Thus

P(B) = P(A) + P(B \A)

The result now follows from the fact that P(B \ A)= 0.

Theorem15.3.4 If $A, B \in \mathrm{F}$, $P(A \cap B^c) = P(A) - P(A \cap B)$

Proof:

We have

$(A \cap B^c) \cup (A \cap B) = A$

$\therefore P[(A \cap B^c) \cup (A \cap B)] = P(A)$

$\Rightarrow P(A \cap B^c) + P(A \cap B) = P(A)$

$\Rightarrow P(A \cap B^c) = P(A) - P(A \cap B)$

We can similarly show that

$$P(A^c \cap B) = P(B) - P(A \cap B)$$

Theorem15.3.5. If $A, B \in \mathrm{F}$, $P(AUB) = P(A) + P(B) - P(A \cap B)$

Proof:

We have

$$A \cup B = (A^c \cap B) \cup (A \cap B) \cup (A \cap B^c)$$

$$\therefore P(A \cup B) = P[(A^c \cap B) \cup (A \cap B) \cup (A \cap B^c)]$$

$$= P(A^c \cap B) + P(A \cap B) + P(A \cap B^c)$$

$$= P(B) - P(A \cap B) + P(A \cap B) + P(A) - P(A \cap B)$$

$$= P(B) + P(A) - P(A \cap B)$$

Theorem 15.3.6: If A and B are any two events, then

$$P(A \setminus B) = P(A) - P(A \cap B)$$

Proof. Now A can be decomposed into the mutually exclusive events A \ B and A n B; that is, $A = (A \setminus B) \cup (A \cap B)$. Thus by [P3],

$$P(A) = P(A \setminus B) + P(A \cap B)$$

from which our result follows.

15.4 Simple Illustrations For Three Events

Applying the above theorem 2.3.5 and 2.3.6 twice we obtain

Theorem 15.4.1: For any events A, B and C,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Let $D = B \cup C$. Then $A \cap D = A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and

$$P(A \cap D) = P(A \cap B) + P(A \cap C) - P(A \cap B \cap A \cap C) = P(A \cap B) + P(A \cap C) - P(A \cap B \cap C)$$

Thus

$$P(A \cup B \cup C) = P(A \cup D) = P(A) + P(D) - P(A \cap D)$$

$$= P(A) + P(B) + P(C) - P(B \cap C) - [P(A \cap B) + P(A \cap C) - P(A \cap B \cap C)]$$

$$= P(A) + P(B) + P(C) - P(B \cap C) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C)$$

Example 15.4.1: Three horses A, B and C are in a race; A is twice as likely to win as B and B is twice as likely to win as C. What are their respective probabilities of winning,

i.e. P(A), P(B) and P(C)?

Let $P(C) = p$; since B is twice as likely to win as C, $P(B) = 2p$; and since A is twice as likely to win as B, $P(A) = 2P(B) = 2(2p) = 4p$. Now the sum of the probabilities must be 1; hence

$$p + 2p + 4p = 1 \quad \text{or} \quad 7p = 1 \quad \text{or } p = 1/7$$

Accordingly,

$$P(A) = 4p = 4/7,\ P(B) = 2p = 2/7,\ P(C) = p = 1/7$$

Question: What is the probability that B or C wins, i.e. P({B ∪ C})? By definition

P({B ∪ C)) =P(B)+P(C)=2/7+1/7=3/7

15.5. An infinite sample space

Exercise 2.5.1. Consider the sample space $\{a_1, a_2, a_3, a_4, a_5, a_6\}$ of some experiment, where outcome a_i has probability p_i for i = 1, . . . , 6. We perform this experiment twice in such a way that the associated probabilities are

$$P((a_i, a_i)) = p_i, \quad \text{and} \quad P((a_i, a_j)) = 0 \text{ if } i = j, \quad \text{for } i, j = 1, \ldots, 6.$$

Check that P is a probability function on the sample space ? = $\{a_1, \ldots, a_6\} \times \{a_1, \ldots, a_6\}$ of the combined experiment. What is the relationship between the first experiment and the second experiment that is determined by this probability function?

We started this section with the experiment of throwing a coin twice. If we want to learn more about the randomness associated with a particular experiment, then we should repeat it more often, say n times. For example, if we perform an experiment with outcomes 1 (success) and 0 (failure) five times, and we consider the event A "exactly one experiment was a success," then this event is given by the set

A = {(0, 0, 0, 0, 1), (0, 0, 0, 1, 0), (0, 0, 1, 0, 0), (0, 1, 0, 0, 0), (1, 0, 0, 0, 0)} in A = {0, 1} × {0, 1} × {0, 1} × {0, 1} × {0, 1}. Moreover, if success has probability p and failure probability 1 - p, then

$$P(A) = 5 \cdot (1 - p)^4 \cdot p,$$

since there are five outcomes in the event A, each having probability $(1 - p)^4 \cdot p$.

Exercise 15.5.2. What is the probability of the event B "exactly two experiments were successful"?

In general, when we perform an experiment n times, then the corresponding sample space is

$$\omega = \omega_1 \times \omega_2 \times \cdots \times \omega_n,$$

where ω_i for i = 1, . . . , n is a copy of the sample space of the original experiment. Moreover, we assign probabilities to the outcomes $(\omega_1, \ldots, \omega_n)$ in the standard way described earlier, i.e.,

$$P((\omega_1, \omega_2, \ldots, \omega_n)) = p_1 \cdot p_2 \cdots p_n, \quad \text{if each } \omega_i \text{ has probability } p_i.$$

We require some rules to assign probabilities to some basic events in $\mathcal{F}$.. For other events we can compute the probabilities in terms of the probabilities of these basic events.

15.6. Probability assignment in a discrete sample space

Consider a finite sample space $S = \{ s_1, s_2, ...s_n \}$. Then the sigma algebra $\mathcal{F}$ is defined by the power set of S. For any e*lementary event* $\{s_i\} \in \mathcal{F}$, we can assign a probability $P(\{s_i\})$ such that,

$$\sum_{i=1}^{N} P(\{s_i\}) = 1$$

For any event $A \in \mathcal{F}$, we can define the probability

$$P(A) = \sum_{A_i \in A} P(\{ A_i \})$$

In a special case, when the outcomes are equi-probable, we can assign equal probability p to each elementary event.

$$\therefore \sum_{i=1}^{n} p = 1$$
$$\Rightarrow p = 1/n$$

$$\therefore P(A) = P\left(\bigcup_{S_i \in A} \{S_i\} \right)$$
$$= n(A) \frac{1}{n} = \frac{n(A)}{n}$$

Example15.6.1 Consider the experiment of rolling a fair die considered in earlier example.

Suppose $A_i, i = 1,..,6$ represent the elementary events. Thus A_1 is the event of getting '1', A_2 is the event of getting '2' and so on.

Since all six disjoint events are equiprobable and $S = A_1 \cup A_2 \cup .... \cup A_6$ we get

$$P(A_1) = P(A_2) = ... = P(A_6) = \frac{1}{6}$$

Suppose A is the event of getting an odd face. Then

$$A = A_1 \cup A_3 \cup A_5$$
$$\therefore P(A) = P(A_1) + P(A_3) + P(A_5) = 3 \times \frac{1}{6} = \frac{1}{2}$$

Example 15.6.2. Consider the experiment of tossing a fair coin until a head is obtained discussed in Example 3. Here $S = \{H, \ TH, TTH, ...\}$. Let us call

$$s_1 = H$$
$$s_2 = TH$$
$$s_3 = TTH$$

and so on. If we assign, $P(\{s_n\}) = \frac{1}{2^n}$ then $\sum_{s_n \in S} P(\{s_n\}) = 1$. Let $A = \{s_1, s_2, s_3\}$ is the event of obtaining the head before the 4th toss. Then

$$P(A) = P(\{s_1\}) + P(\{s_2\}) + P(\{s_3\})$$
$$= \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} = \frac{7}{8}$$

Probability assignment in a continuous space

Suppose the sample space S is continuous and uncountable. Such a sample space arises when the outcomes of an experiment are numbers. For example, such sample space occurs when the experiment consists in measuring the voltage, the current or the resistance. In such a case, the sigma algebra consists of the Borel sets on the real line.

Suppose $S = R$ and $f : R = R$ is a non-negative integrable function such that,

$$\int_A f(x)\, dx$$

For any Borel set A,

$P(A) = \int_A f(x)\, dx$ defines the probability on the Borel sigma-algebra B.

We can similarly define probability on the continuous space of $R^2 = R^3$ etc.

Example 15.6.3. Suppose

$$f_X(x) = \begin{cases} \dfrac{1}{b - a} & \text{for } x \in [a,\ b] \\ 0 & \text{otherwise} \end{cases}$$

Then for $[a_1,\ b_1] \subseteq [a,\ b]$

$$P([a_1,\ b_1]) = \frac{b_1 - a_1}{b - a}$$

Example 15.6.4. Consider $S = R^2$ the two-dimensional Euclidean space.

Let $S_1 \subseteq R^2$ and $|S_1|$ represents the area under S_1.

$$f_X(x) = \begin{cases} \dfrac{1}{|S_1|} & \text{for } x \in S_1 \\ 0 & \text{otherwise} \end{cases}$$

Then for any $A \subseteq S_1$,

$$P(A) = \frac{|A|}{|S_1|}$$

This example interprets the geometrical definition of probability.

15.7 Lesson End Excercises

1. Let A and B be events with P(A)=0.4,P(B)=0.2.Find P(A n B), P(A^c n B^c),P(A^c U B^c) and P(B n A^c).

2. State the limitations of classical definition of Probability.

15.8. Suggested Readings

1. Goon,Gupta;Das Gupta (1991).Fundamental of Statistics
2. Gupta,S.C. and Kapoor,V.K. Fundamental of Mathematical Statistics
3. Hoel,P.G. (1971).Introductory of Mathematical Statistics
4. Hogg R.V and Craig,A.T.Introduction to Mathematical Statistics
5. Hogg ,RV and Tanis,EA(1993).Probability and Statistical Inference
6. Mood,AM,Bose DC and Graybill F A.Introduction to the Theory of Statistics
7. Rohtagi,VK.An Introduction to Probability Theory and Mathematical Statistics

****

UNIT-1V **LESSON-16**

CONDITIONAL PROBABILITY AND INDEPENDENT EVENTS

STRUCTURE

16.1 Objectives: The objectives of this lesson are

1. To familiarise students to the concept of conditional probability
2. To introduce students the concept of independence of events
3. To prove some important probability laws

16.2 Introduction

The mathematical theory of probability has its roots in attempts to analyze games of chance by Gerolamo Cardano in the sixteenth century, and by Pierre de Fermat and Blaise Pascal in the seventeenth century (for example the “problem of points”). Christiaan Huygens published a book on the subject in 1657.

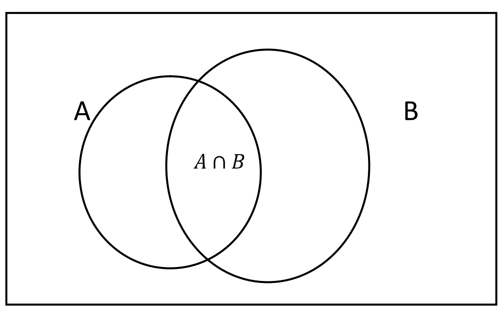
Initially, probability theory mainly considered **discrete** events, and its methods were mainly combinatorial. Eventually, analytical considerations compelled the incorporation of **continuous** variables into the theory.

This culminated in modern probability theory, on foundations laid by Andrey Nikolaevich Kolmogorov. Kolmogorov combined the notion of sample space, introduced by Richard von Mises, and **measure theory** and presented his axiom system for probability theory in 1933. Fairly quickly this became the mostly undisputed axiomatic basis for modern probability theory but alternatives exist, in particular the adoption of finite rather than countable additivity by Bruno de Finetti.

16.3 Conditional Probability:

Let B be an arbitrary event in a sample space S with probability P(B)>0.The probability that an event A occurs once B has occurred ,or, in other words ,the conditional probability of A given B denoted by P(A|B),is defined as

$$P(A \mid B)\frac{P(A \cap B)}{P(B)}, P(B) > 0$$



Reduced sample space

In a certain sense ,P(A|B) measures the relative probability of A w.r.t. the new sample space (the reduced space of B) If S is an equiprobable space, then

$$P(A|B)=\frac{n(A\cap B)}{n(S)},P(B)\frac{n(B)}{n(S)}$$

$$P(A|B)=\frac{n(A\cap B)}{n(B)}=\frac{\text{number of ways both A and B can occur}}{\text{number of ways B can occur}}$$

Similary,we define

$$P(B|A)=\frac{P(A\cap B)}{P(A)},P(A)>0$$

Example: A family has two children. Find the probability that both children are boys if it is known that (i) one of the children is a boy,(ii) the older child is a boy.

Sol. Here S={bb, bg, gb, gg}.The reduced space is

(i) B={bb,bg,gb},so required probability p =1/3

(ii) B={bb,bg}, so required probability p =2/4 =1/2

Example:

The question, "Do you smoke?" was asked of 100 people. Results are shown in the table.

	Yes	**No**	**Total**
Male	19	41	**60**
Female	12	28	**40**
Total	**31**	**69**	**100**

- What is the probability of a randomly selected individual being a male who smokes? This is just a joint probability. The number of "Male and Smoke" divided by the total = 19/100 = 0.19
- What is the probability of a randomly selected individual being a male? This is the total for male divided by the total = 60/100 = 0.60. Since no mention is made of smoking or not smoking, it includes all the cases.
- What is the probability of a randomly selected individual smoking? Again, since no mention is made of gender, this is a marginal probability, the total who smoke divided by the total = 31/100 = 0.31.
- What is the probability of a randomly selected male smoking? This time, you're told that you have a male - think of stratified sampling. What is the probability

that the male smokes? Well, 19 males smoke out of 60 males, so 19/60 = 0.31666...

- What is the probability that a randomly selected smoker is male? This time, you're told that you have a smoker and asked to find the probability that the smoker is also male. There are 19 male smokers out of 31 total smokers, so 19/31 = 0.6129 (approx)

16.4 Conditional probability as a probability measure:

To show that that conditional probability function P(A|B) satisfies the three axioms of probability theory.

(P_1) $(P_1)(0 \leq P(A \mid B) \leq 1$

(P_2) $P(S \mid B) = 1$

(P_3) $\{A_1 \cup A_2 \cup ....\mid B\} = P(A_1|B) - P(A_2|B)....$

Where A_1, A_2............ are mutually exclusive/events

Proof: (1) Since $A \cap B \subset B$, we get $P(A \cap B) \leq P(B)$. Thus

$P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} \leq 1$, and is also non negative because $P(A \cap B) \geq 0$ *and* $P(B) \geq 0$

(2) Since $S \cap B = B$, $P(S|B) = \dfrac{P(S \cap B)}{P(B)} = \dfrac{P(B)}{P(B)} = 1$

(3) If $A_1 \cap A_2 = \phi\,(i \neq j)$ then $(A_i \cap B) \cap (A_j \cap B) = \phi$, so $(A_i \cap B)$ and $(A_j \cap B)$ are disjoint events. Now

$$\frac{P\{(A_1 \cup A_2 \cup ....) \cap B\}}{P(B)} = \frac{P(A_1 \cap B)}{P(B)} + \frac{P(A_2 \cap B)}{P(B)} + ...$$

16.5 Compound Probability theorem or Multiplication theorem of Probability:

If A and B are two possible events of a sample space S, then

$P(A\cap B) = P(A)P(B|A) = P(B)P(A|B)$

Proof: Denote the number of cases favourable to the events $A \cap B, \overline{A} \cap B, A \cap \overline{B}$ and $\overline{A} \cap \overline{B}$ by n_{11}, n_{12}, n_{21}, n_{22} respectively, as shown below

Events	A	$\bar{A}$	Totals
B	n_{11}	n_{12}	$n_{11} + n_{12}$
$\bar{B}$	n_{21}	n_{22}	$n_{21} + n_{22}$
Totals	$n_{11} + n_{21}$	$n_{12+} n_{22}$	N

Let N be the total numbers of cases in the sample space .Then by classical definition of probability,

$$P(A) = \frac{n_{11} + n_{21}}{n_{11}}, P(B) = \frac{n_{11} + n_{12}}{N}, P(A \cap B) = \frac{n_{11}}{N}, P(A | B) = \frac{n_{11}}{n_{11} + n_{12}}$$

Here P(A|B) =probability of event A relative to the space of event B

Obviously,

$$P(A | B) = \frac{n_{11}}{n_{11} + n_{12}} = \frac{n_{11}/N}{n_{11} + n_{22}/N} = \frac{P(A \cap B)}{P(B)}, P(B) > 0$$

Similarly, we have

$$P(B | A) = \frac{n_{11}}{n_{11} + n_{12}} = \frac{n_{11}/N}{n_{11} + n_{22}/N} = \frac{P(A \cap B)}{P(B)}, P(A) > 0$$

16.6 Extension to n events:

For three possible events A,B and C

$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$

$P(A \cap B \cap C)$ $= P[(A \cap B)C] = P(A \cap B)P(C| A \cap B)$

$= P(A)P(B|A)P(C| A \cap B)$

In general for n events $A_1, A_2 \ldots\ldots\ldots\ldots A_n$

$P(A_1 \cap A_2 \cap \ldots\ldots\ldots\ldots \cap A_n)$ =

$P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2)\ldots\ldots\ldots\ldots P(A_n | A_1 \cap A_2 \cap \ldots\ldots\ldots \cap A_{n-1})$

Proof: $P(A_1 \cap A_2 \cap \ldots\ldots\ldots \cap A_n) = P(A_1).\dfrac{P(A_1 \cap A_2)P(A_1 \cap A_2 \cap A_3)}{P(A_1)\ .\ (A_1 \cap A_2)}$ **..........**

$$\frac{P(A_1 \cap A_2 \cap \ldots\ldots\ldots\ldots \cap A_{n-1})\, P(A_1 \cap A_2 \cap \ldots\ldots\ldots\ldots \cap A_n)}{P(A_1 \cap A_2 \cap \ldots\ldots\ldots\ldots \cap A_{n-2})\, P(A_1 \cap A_2 \cap \ldots\ldots\ldots\ldots \cap A_{n-1})}$$

$= P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2)\ldots\ldots\ldots\ldots P(A_n | A_1 \cap A_2 \cap \ldots\ldots\ldots \cap A_n)$

[by def. of conditional probability]

16.7 Some Important Theorems:

Theorem 3.7.1:

For any three events A,B and C

$P(A \cup B | C)$ = $P(A|C)+P(B|C) - P(A \cup B | C)$

Proof: We have,

$P(A \cup B) = P(A)+P(B) - P(A \cap B)$

$\Rightarrow P[(A \cap C) \cup (B \cap C)] = P(A \cap C) + (B \cap C) - P(A \cap B \cup C)$

Dividing both sides by P(C) > 0, we get

$$\frac{P[(A\cap C)\cup(B\cap C)]}{P(C)}=\frac{P(A\cap C)+P(B\cap C)-P(A\cap B\cap C)}{P(C)}$$

$$=\frac{P(A\cap C)}{P(C)}+\frac{P(B\cap C)}{P(C)}-\frac{P(A\cap B\cap C)}{P(C)}$$

$$\Rightarrow\frac{P[(A\cap B)\cap C]}{P(C)}\ =\ P(A\mid C)+P(B\mid C)-P(A\cap B\mid C)$$

$$\Rightarrow P[(A\cup B)\mid C]=P(A\mid C)+P(B\mid C)-P(A\cap B\mid C)$$

Theorem 16.7.2:

For any three events A, B and C

$$P(A\cap\overline{B}\mid C)+P(A\cap B\mid C)=P(A\mid C)$$

Proof: $P(A\cap\overline{B}\mid C)+P(A\cap B\mid C)=\frac{P(A\cap\overline{B}\cap C)}{P(C)}+\frac{P(A\cap B\cap C)}{P(C)}$

$$=\frac{P(A\cap\overline{B}\cap C)+P(A\cap B\cap C)}{P(C)}=\frac{P(A\cap C)}{P(C)}=P(A\mid C)$$

Theorem16.7.3:

For any three events A,B and C in the sample space S such that $B\subset C$ and P(A)>0,we have

$P(B|A)\le P(C|A)$

P (C∩ A)

Proof: $P(C\mid A)=\frac{P(A\cup C)}{P(A)}=\frac{P(A\cap\overline{B}\cap C)\cup P(A\cap\overline{B}\cap C)}{P(C)}$

P(

$$=\frac{P(A\cap\overline{B}\cap C)+P(A\cap B\cap C)}{P(A)}=P\overline{B}\cap C\mid A)+P(B\cap C\mid A)$$

But $B \subset C \Rightarrow B \cap C = B$

Hence

$$P(C \mid A) = P(\overline{B} \cap C \mid A) + P(B \mid A)$$

$$\Rightarrow P(C \mid A) \geq P(B \mid A)$$

16.8 Independence of Events:

Two events A and B are said to be independent if

$$P(A \cap B) = P(A)P(B)$$

In other words Two events A and B are independent if

$P(A|B) = P(A)$ or $P(B|A) = P(B)$

Three events A,B and C are independent if

$P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, $P(B \cap C) = P(B)\,P(C)$ and $P(A \cap B \cap C) = P(A)P(B)P(C)$

A family of events is completely independent iff its every finite subcollection of events is independent.

In general events $A_1, A_2 .... A_n$ are said to be pair wise independent if

$$P(A_i \cap A_j) = P(A_i)P(A_j), \forall i, j, i = j$$

Events $A_1, A_2 .... A_n$ are said to be mutually independent if

$$P(A_i \cap A_j) = P(A_i)P(A_j), \forall i, j, i \neq j$$

$$P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)\,P(A_k), \forall i, j, i \neq j \neq k$$

....

$$(.)P(A_1 \cap A_2 \cap .... \cap A_n) = P(A_1)P(A_2)\,P(A_3) .... P(A_n)$$

16.9 Conditions for Mutual Independence of n Events:

The events in the sample space S are said to be mutually independent if the probability of simultaneous occurrence of finite number of events is equal to the product of their separate probabilities i.e.

Events A_1, A_2...............A_n are said to be mutually independent if

(i) $P(A_i \cap A_j) = P(A_i)P(A_j), \forall i, j, i \neq j$

(ii) $P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)\,P(A_k), \forall i, j, i \neq j \neq k$

....

$(.)P(A_1 \cap A_2 \cap .... \cap A_n) = P(A_1)P(A_2)\,P(A_3)..........P(A_n)$

It is to be noted here that above equations give respectively $\binom{n}{2}, \binom{n}{3} .... \binom{n}{n}$ conditions to be satisfied by A_1, A_2...............A_n. Hence the total number of conditions for mutual independence of A_1, A_2...............A_n is

$$\binom{n}{2}, \binom{n}{3} .... \binom{n}{n} = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \binom{n}{3} + ... \binom{n}{n} - \binom{n}{0} - \binom{n}{1}$$

$= 2^n - n - 1$

3.10 Independence versus Mutual Exclusiveness:

(i) If A and B are mutually exclusive events and P(A).P(B)>0, then they cannot be independent.

(ii) If A and B are independent and P(A).P(B)>0, then they cannot be mutually exclusive.

Proof: We are given that $P(A) \neq 0$ and $P(B) \neq 0$, i.e. P(A).P(B)>0

(i) If the events A and B are mutually disjoint then

$A \cap B = \phi$ so $P(A \cap B) = 0 \neq P(A)P(B)$

Hence A and B cannot be independent.

(ii) If the events A and B are independent ,then $P(A \cap B) = P(A).P(B) > 0$

Thus $P(A \cap B) \neq 0$. So A and B cannot be mutually disjoint.

16.11 Some Important Theorems:

Theorem 16.11.1:

If A and B are independent events, then

(i) A and $\overline{B}$ are independent (ii) $\overline{A}$ and B are independent (iii) $\overline{A}$ and $\overline{B}$ are independent

Proof: Since A and B are independent so we have $P(A \cap B) = P(A)P(B)$ (I)

To prove that A and $\overline{B}$ are independent we are to prove that $P(A \cap \overline{B}) = P(A)P(\overline{B})$

We know that $P(A \cap \overline{B}) = P(A) - P(A \cap B)$

$= P(A) - P(A)\,P(B)$ [Using (I)]

$P(A \cap \overline{B}) = P(A)[1-P(B)] = P(A)\,P(\overline{B})$

Thus A and $\overline{B}$ are independent.

(ii) To prove that $\overline{A}$ and B are independent we are to prove that $P(\overline{A} \cap B) = P(\overline{A})P(B)$

We have $P(\overline{A} \cap B) = P(B) - P(A \cap B)$

$= P(B) - P(A)P(B)$ [Using (I)]

So $P(\overline{A} \cap B) = P(B) - P(A)\,P(B)$

$= P(B)[1-P(A)] = P(\overline{A})P(B)$

Thus $\overline{A}$ and B are independent.

(iii) To prove that $\overline{A}$ and $\overline{B}$ are independent we are to prove that $P(\overline{A} \cap \overline{B}) = P(\overline{A})P(\overline{B})$

By DeMorgan's law we have

$P(\overline{A} \cap \overline{B}) = 1 - P(A \cup B) = 1 - [P(A) + P(B) - P(A \cap B)]$

$= 1 - P(A) - P(B) + P(A)\,P(B)$

$= [1-P(A)][1-P(B)] = P(\overline{A})\,P(\overline{B})$

Hence $\overline{A}$ and $\overline{B}$ are independent.

Theorem16.11. 2:

If A,B and C are mutually events then $A \cup B$ and C are also independent.

Proof: We are to prove that

$P[(A \cup B) \cap C] = P(A \cup B)\, P(C)$

Consider $P[(A \cup B) \cap C] = P[(A \cap C) \cup (B \cap C)]$ [by distributive law]

$= P(A \cap C) + P(B \cap C) - P(A \cap B \cap C)$

$= P(A)P(C) + P(B)P(C) - P(A)P(B)P(C)$ [Since A,B,C are independent]

$= P(C)[P(A) + P(B) - P(A)P(B)]$

$= P(C)[P(A) + P(B) - P(A \cap B)]$

$= P(C)[P(A \cup B]$

Hence $A \cup B$ and C are independent.

16.12 Illustrations:

1) There is a 10% chance that you will travel to Mexico and meet a tall dark stranger, and a 90% chance that you will meet a tall dark stranger. Find the he probability that you will travel to Mexico given that you meet a tall dark stranger is:

Sol. Define the following events:

A: You travel to Mexico

B: you meet a tall dark stranger

Then we are given

$P(A \cap B)$ = P[you travel to Mexico and meet a tall dark stranger] =.10

P(B) = P[you meet a tall dark stranger] =.90

We are to find P[that you travel to Mexico given that you meet a tall dark stranger] i.e.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{.10}{.90} = \frac{1}{9}$$

2) Given that P(A) =1/3, P(B)=3/4, $P(A \cup B)$=11/12. Fnd P(A|B) and P(B|A)

Sol. We know that

$$P(A \cap B) = P(A) + P(B) - P(A \cup B)$$

$$= \frac{1}{3} + \frac{3}{4} - \frac{11}{12} = \frac{1}{6}$$

Hence

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad \frac{1}{6} \Big/ \frac{3}{4} = 2/9$$

Similarly,

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \quad \frac{1}{6} \Big/ \frac{1}{3} = 1/2$$

3) A box contains 10 gold coins and 8 silver coins. Two successive drawings of 4 coins are made such that (i) coins are replaced before the second trial(ii) the coins are not replaced before the second trial. Find the probability that the first drawing will give 4 gold and the second 4 silver coins.

Sol: Let a denote the event of drawing 4 gold coins in the first draw and B denote the event of drawing 4 silver coins in the second draw.Then we are to find P(A∩B).

(i) **Draws with replacement**: If the coins drawn in the first draw are replaced back in the box before the second draw then the events A and B are independent and required probability is given by

$$P(A \cap B) = P(A)P(B)$$

First draw: for the first draw, the probability of drawing 4 gold coins in first draw is

$$P(A) = \frac{\binom{10}{4}}{\binom{18}{4}}$$

Since out of total number of 18 coins 4 can be drawn in $\binom{18}{4}$ ways which gives the exhaustive number of cases and favourable number of cases is $\binom{10}{4}$ as out of 10 gold coins 4 can be drawn in $\binom{10}{4}$ ways

Second draw: When the coins in the first draw are replaced before the second draw the box contains again 18 coins hence probability of drawing 4 silver coins in second draw is

$$P(B) = \frac{\binom{8}{4}}{\binom{18}{4}}$$

Hence required probability , $P(A \cap B) = P(A)P(B)$

$$= \frac{\binom{10}{4}}{\binom{18}{4}} \times \frac{\binom{8}{4}}{\binom{18}{4}}$$

(ii) **Draws without replacement**: If the coins drawn are not replaced back before the second draw, then the events A and B are not independent and the required probability is given by

$$P(A \cap B) = P(A)P(B|A)$$

Again, $P(A) = \frac{\binom{10}{4}}{\binom{18}{4}}$

Now if 4 gold coins which were drawn in the first draw are not replaced back then there remain 14 coins in box and P(B|A) is the probability of drawing 4 silver coins from the box containing14 coins out of which 6 are gold coins and 8 are silver coins.

Hence $P(B \mid A) = \dfrac{\binom{8}{4}}{\binom{14}{4}}$

Therefore required probability is

$$P(A \cap B) = P(A)P(B|A) = \frac{\binom{10}{4}}{\binom{18}{4}} \text{ x } \frac{\binom{8}{4}}{\binom{14}{4}}$$

4) Suppose an urn contains7 black balls and 5 white balls. Two balls are drawn from the urn without replacement. Assuming that each ball in the urn is equally likely to be drawn, find the probability that both the balls drawn are black.

Sol. Let A and B be the events that first and second balls drawn are black. Given the first ball selected is black, there remain six black balls and five white balls in the urn. So

P(A)= 7/12 and P(B|A) = 6/11

Required probability = $P(A \cap B) = P(A)P(B|A)$ = 7/12 x 6/11 =42/132

16.13 SUMMARY:

In this chapter we discussed conational probability, independence of events ,various probability laws like addition theorem of probability, multiplication theorem of probability and some other important theorems. For more clear understanding of the students numerical examples have been taken.

16.14 SELF ASSESSMENT:

1. If P(A) = a,P(B)= b,then show that

$$P(A \mid B) \geq \frac{a+b-1}{b}$$

2. A bag contains 15 items, of which 4 are defectives. The items are selected at random one by one and examined. The ones examined are not put back .What is the chance that the 10^{th} one examined is the last defective?

3. A student is to appear for two tests in which his respective chances of winning are 0.5 and 0.7 and loosing both the test is 0.2 .Find the probability that the student will win test-2 when he has already won test-1.

16.15 FURTHER READINGS:

1. Theory of Probability, P.Mukhopadhyay

2. An Introduction to Theory of Probability and Mathematical Statistics : V K Rohatgi

3. An Introduction to Probability Theory and its Application,w.Feller

4. Fundamentals of Mathematical Statistics : S C Gupta and V K Kapoor

5. New Mathematical Statistics : Bansi Lal and Sanjay Arora

****

UNIT-IV **LESSON-17**

BAYES THEOREM AND ITS APPLICATIONS

STRUCTURE

17.1 OBJECTIVES: The objectives hare are

1. To give an overview of Bayes theorem
2. To discuss significance and applicability of Bayes theorem

17.2 INTRODUCTION

In probability theory and applications, **Bayes' theorem** shows how to determine *inverse probabilities*: knowing the conditional probability of *B* given *A*, what is the conditional probability of *A* given *B*? This can be done, but also involves the so-called prior or unconditional probabilities of *A* and *B*.

This theorem is named for Thomas Bayes (pronounced as "bays") and often called **Bayes' law** or **Bayes' rule**. Bayes' theorem expresses the conditional probability, or "posterior probability", of a event in terms of the " a priori probabilities" and the conditional probabilities.

17.3 BAYES THEOREM:

Statement: Let E_1, E_2 ....E_n are mutually disjoint events in a sample space

'S' with $P(E_i) \neq 0$ (i=1,2,…….n) then for any arbitrary event A which is a subsets of $U_{i=1}^{n} E_i$ such that P(A)>0,we have

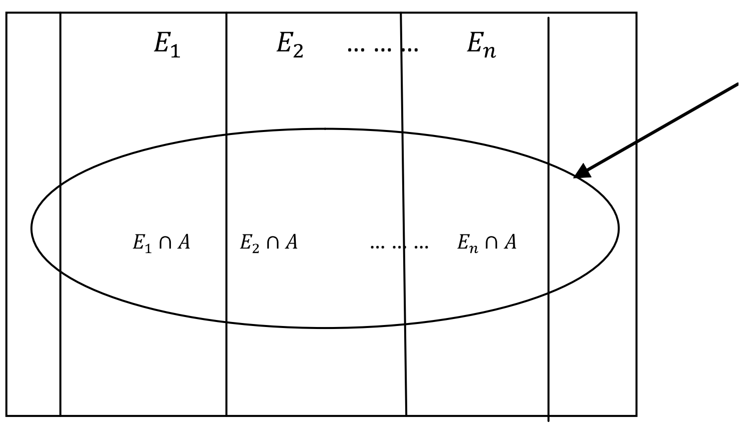
$$P(E_i \mid A) = \frac{P(E_i)P(A \mid E_i)}{\sum_{i=1}^{n} P(E_i)P(A \mid E_i)}$$

Proof: Let S be the sample space and $E_1, E_2, ...E_n$ are mutually disjoints in the sample space S.Since $A \subset U_{i=1}^{n} E_i$, So we have

$$A = A \cap (U_{i=1}^{n} E_i) = U_{i=1}^{n} (A \cap E_i) \qquad \text{(By distributive law)}$$

Since $(A \cap E_i) \subset E_i (i = 1, 2, ...n)$ are mutually disjoint events, so by addition theorem of probability

$$P(A) = P[U_{i=1}^{n} (A \cap E_i)] = \sum_{i=1}^{n} P(A \cap E_i) = \sum_{i=1}^{n} P(E_i)P(A \mid E_i) \qquad (4.3.1)$$



Also by compound probability theorem, we have

$$(A \cap E_i) = P(A)\,P(E_i \mid A) \qquad (4.3.2)$$

By conditional probability

$$P(E_i \mid A) = \frac{P(A \cap E_i)}{P(A)} = \frac{P(E_i)P(A \mid E_i)}{\sum_{i=1}^{n} P(E_i)P(A \mid E_i)} \qquad \text{[Using (4.3.1) and (4.3.2)]}$$

Remarks:

(1) The probabilities $P(E_1), P(E_2), \ldots\ldots\ldots\ldots, P(E_n)$ are known as a priori probabilities because they exist before we gain any information from the experiment itself.

(2) The probabilities $P(A \mid E_i)$, (i=1,2,…….n) are called likelihoods because they indicate how likely the event A under consideration is to occur, given each and every a priori probabilities.

(3) The probabilities $P(E_i \mid A)$, (i=1,2,…….n) are called posterior probabilities because they are determined after the results of the experiment are known.

17.4 BAYES THEOREM FOR FUTURE EVENTS:

Statement:Let $E_1, E_2, \ldots\ldots\ldots\ldots E_n$ are mutually disjoint events in a sample space 'S' with $P(E_i) \neq 0$ (i=1,2,…….n) then for any arbitrary event A which is a subsets of $U_{i=1}^{n} E_i$ with P(A)>0 that is conditioned on the occurrence of the event C with P(C) >0 and $C\ U_{i=1}^{n} E_i$ then

$$P(C \mid A) = \frac{\sum_{i=1}^{n} P(E_i)P(A \mid E_i)P(C \mid E_i \cap A)}{\sum_{i=1}^{n} P(E_i)P(A \mid E_i)}$$

Proof : Since the occurrence of the event A implies the occurrence of one and only one of the event $E_1, E_2, \ldots\ldots\ldots\ldots E_n$, the event C (given that A has occurred)can occur in the following mutually exclusive ways:

$C \cap E_1, C \cap E_2 \ldots\ldots\ldots\ldots, C \cap E_n$ i.e.

$C = (C \cap E_1) \cap (C \cap E_1) \cap, \ldots\ldots\ldots\ldots \cap (C \cap E_1)$

$\Rightarrow (C \mid A) = [(C \cap E_1 \mid A)] \cap [(C \cap E_2 \mid A)] \cap \ldots\ldots\ldots\ldots \cap [(C \cap E_2 \mid A)]$

So

$P(C \mid A) = P[(C \cap E_1 \mid A)] + P[(C \cap E_2 \mid A)] + \ldots\ldots\ldots\ldots + P[(C \cap E_2 \mid A)]$

$$= \sum_{i=1}^{n} P[(C \cap E_1 \mid A)]$$

$$= \sum_{i=1}^{n} P(E_i \mid A) P[C \mid E_i \cap A] \qquad (4.4.1)$$

But

$$P(E_i \mid A) = \frac{P(A \cap E_i)}{P(A)} = \frac{P(E_i)P(A \mid E_i)}{\sum_{i=1}^{n} P(E_i)P(A \mid E_i)} \qquad \text{[Using Bayes theorem]}$$

Substituting this value in (4.4.1), we get

$$P(C \mid A) = \frac{\sum_{i=1}^{n} P(E_i)P(A \mid E_i)P(C \mid E_i \cap A)}{\sum_{i=1}^{n} P(E_i)P(A \mid E_i)}$$

17.5 Some Illustrations on Bayes Theorem:

1) Urn I contains 1 white ,2 black and 3 red balls, Urn II contains 2 white ,1 black and 1 red balls and Urn III contains 4 white ,5 black and 3 red balls. One urn is chosen at random and two balls are drawn .They happen to be white and red. Find the probability that they came from urn I,II or III.

Sol.Let E_1, E_2 and E_n denote the events that the urn I,II and III is selected respectively. Let A be the event that the two balls taken from the selected urn are white and red. Then $P(E_1) = P(E_2) = P(E_3) = 1/3$

And

$$P(A \mid E_1) = \frac{1x3}{\binom{6}{2}} = 1/5, \qquad P(A \mid E_2) = \frac{2x1}{\binom{4}{2}} = 1/3 \quad \text{and } P(A \mid E_3) = \frac{4x3}{\binom{12}{2}} = 2/11$$

Now

$$P(E_1 \mid A) = \frac{P(E_1)P(A \mid E_1)}{\sum_{i=1}^{n} P(E_i)P(A \mid E_i)} = \frac{\frac{1}{3}}{\frac{1}{3} x \frac{1}{5} x \frac{1}{3} x \frac{1}{3} x \frac{1}{3} x \frac{2}{11}} = \frac{33}{118}$$

Similarly ,

$$P(E_2 \mid A) = \frac{P(E_2)P(A \mid E_2)}{\sum_{i=1}^{3} P(E_i)P(A \mid E_i)} = \frac{55}{118}$$

and $P(E_3 \mid A) = \frac{P(E_3)P(A \mid E_3)}{\sum_{i=1}^{3} P(E_i)P(A \mid E_i)} = \frac{30}{118}$

2) In answering a question on a multiple choice test a student either knows the answer or he guesses. If p be the probability that he knows the answer and 1-p is the probability that he guesses. Assume that a student who guesses the answer will be correct with probability 1/m ,where m is the number of multiple –choice alternatives. Find the probability that a student knew the answer to a question given that he answered it correctly.

Sol. Define the events:

E_1 : The student knew the correct answer

E_2 : The student guesses the answer

A: The student gets the correct answer

Then, we have

$P(E_1) = p, \qquad P(E_2) = 1 - p$

P(A| E_1) =P[student gets right answer given that he knew the correct answer] = 1

$P(A \mid E_2) = \frac{1}{m}$

We are to find

P(E_1 |A)

By Bayes theorem

$$P(E_1 \mid A) = \frac{P(E_1)P(A \mid E_1)}{P(E_1)P(A \mid E_1) + P(E_2)P(A \mid E_2)}$$

$$= \frac{p.1}{p.1 + \frac{1(1-p)}{m}} = \frac{mp}{1 + (m-1)P}$$

3) X write a letter to Y and does not receive a reply. Assuming that one letter in n is lost in mails, find the probability that Y received the letter .It is certain that Y would have replied the letter if he had received it.

Sol. Define the events:

E_1 :Y received the letter

E_2 : Y did not receive the letter

A: X did not receive reply fromY

We are to find P(E_1|A)

We are given

$P(E_2) = 1/n$, $P(E_1) = 1 - \frac{1}{n}$

$P(A|E_2) = 1$, $P(A|E_1) = 1/n$

By Bayes theorem required probability

$$P(E_1 \mid A) = \frac{P(E_1)P(A \mid E_1)}{P(E_1)P(A \mid E_1) + P(E_2)P(A \mid E_2)}$$

$$= \frac{\left(1 - \frac{1}{n}\right)\frac{1}{n}}{\frac{1}{n}\left(1 - \frac{1}{n}\right) + 1\frac{1}{n}} = \frac{n-1}{2n-1}$$

4) The probability that a person can hit a target is 3/5 and the probability that another person can hit the same target is 2/5.But the first person can fire 8 shots in the time the second person fires 10 shots. They fire together. What is the probability that the second person shoots the target.

Sol. Define the events:

E_1 : First person shoots the target

E_2 : Second person shoots the target

A: target is hit

From the given data ,we have

$P(A|E_1) = 3/5$, $P(A|E_2) = 2/5$

The ratio of the shots of the first person to to those of second person in the same time is 8/10 i.e. 4/5

So we have

$$P(E_1) = \frac{4}{5}\ P(E_2)$$

Then by Bayes theorem

$$P(E_2 \mid A) = \frac{P(E_2)P(A \mid E_2)}{P(E_1)P(A \mid E_1) + P(E_2)P(A \mid E_2)}$$

$$= \frac{P(E_2)\frac{2}{5}}{\frac{4}{5}P(E_2)\frac{3}{5} + P(E_2)\frac{2}{5}} = \frac{5}{11}$$

5) Marie is getting married tomorrow, at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 10% of the time. What is the probability that it will rain on the day of Marie's wedding?

***Sol**:* The sample space is defined by two mutually-exclusive events - it rains or it does not rain. Additionally, a third event occurs when the weatherman predicts rain. Notation for these events appears below.

E_1: It rains on Marie's wedding.

E_2: It does not rain on Marie's wedding

A: The weatherman predicts rain.

In terms of probabilities, we know the following:

$P(E_1)$ = 5/365 =0.0136985 [It rains 5 days out of the year.]

$P(E_2)$ = 360/365 = 0.9863014 [It does not rain 360 days out of the year.]

$P(A \mid E_1)$ = 0.9 [When it rains, the weatherman predicts rain 90% of the time.]

$P(A \mid E_2)$ = 0.1 [When it does not rain, the weatherman predicts rain 10% of the time.]

We want to know $P(E_1 \mid A)$, the probability it will rain on the day of Marie's wedding, given a forecast for rain by the weatherman. The answer can be determined from

Bayes' theorem, as shown below.

$P(E_1 | A) = P(E_1) P(A | E_1)$

$P(E_1) P(A | E_1) + P(E_2) P(A | E_2)$

$P(E_1 | A) = (0.014)(0.9) / [(0.014)(0.9) + (0.986)(0.1)]$

$P(E_1 | A) = 0.111$

Note the somewhat unintuitive result. Even when the weatherman predicts rain, it only rains only about 11% of the time. Despite the weatherman's gloomy prediction, there is a good chance that Marie will not get rained on at her wedding.

6) In a bolt factory machines X, Y and Z manufacture respectively 25%, 35% and 40% of the total. Of their output 5%, 4%, 2% are defective bolts. A bolt is drawn at random from the product and is found to be defective. What are the probabilities that it was manufactured by machines X, Y and Z?

Sol. Let E_1, E_2, and E_3 denote the events that a bolt selected at random is manufactured by machines X, Y and Z respectively and A denote the event that it is defective. Then we have $P(E_1) = 0.25$, $P(E_2) = 0.35$, $P(E_3) = 0.40$

The probability of drawing a defective bolt manufactured by machine X and equals $P(A|E_1) = 0.05$

Similarly, $P(A|E_2) = 0.04$ and $P(A|E_3) = 0.02$

The probability that a defective bolt selected at random is manufactured by machines X is given by

$$P(E_1 | A) = \frac{P(E_1)P(A | E_1)}{\sum_{i=1}^{3} P(E_i)P(A | E_i)}$$

$$= \frac{0.25 \text{ x } 0.05}{0.25 \text{ x } 0.05 + 0.35 \text{ x } 0.04 + 0.40 \text{ x } 0.02} = \frac{125}{345} = \frac{25}{69}$$

Similarly, $$P(E_2 | A) = \frac{P(E_2)P(A | E_2)}{\sum_{i=1}^{3} P(E_i)P(A | E_i)}$$

$$= \frac{0.35 \text{ x } 0.05}{0.25 \text{ x } 0.05 + 0.35 \text{ x } 0.04 + 0.40 \text{ x } 0.02} = \frac{28}{69}$$

And $$P(E_3 \mid A) = \frac{P(E_3)P(A \mid E_3)}{\sum_{i=1}^{3} P(E_i)P(A \mid E_i)}$$

$$= \frac{0.40 \text{ x } 0.02}{0.25 \text{ x } 0.05 + 0.35 \text{ x } 0.04 + 0.40 \text{ x } 0.02} = \frac{16}{69}$$

17.6 APPLICATIONS & USEFULNESS OF BAYES THEOREM:

Bayes' Theorem is a theorem of probability theory originally stated by the Reverend Thomas Bayes. It can be seen as a way of understanding how the probability that a theory is true is affected by a new piece of evidence. It has been used in a wide variety of contexts, ranging from marine biology to the development of "Bayesian" spam blockers for email systems. In the philosophy of science, it has been used to try to clarify the relationship between theory and evidence. Many insights in the philosophy of science involving confirmation, falsification, the relation between science and pseudosience, and other topics can be made more precise, and sometimes extended or corrected, by using Bayes' Theorem.

The key idea is that the probability of an event A given an event B (e.g., the probability that one has breast cancer given that one has tested positive in a mammogram) depends not only on the relationship between events A and B (i.e., the accuracy of mammograms) but also on the marginal probability (or "simple probability") of occurrence of each event. For instance, if mammograms are known to be 95% accurate, this could be due to 5.0% false positives, 5.0% false negatives (missed cases), or a mix of false positives and false negatives. Bayes' theorem allows one to calculate the conditional probability of having breast cancer, given a positive mammogram, for any of these three cases. The probability of a positive mammogram will be different for each of these cases. In the example at hand, there is a point of great practical importance that is worth noting: if the prevalence of mammograms resulting positive for cancer is, say, 5.0%, then the conditional probability that an individual with a positive result actually does have cancer is rather small, since the marginal probability of this type of cancer is closer to 1.0%. The probability of a positive result is therefore five times more likely than the probability of the cancer itself. Also, one can deduce that the conditional probability that positive mammogram result implies breast cancer is at most 20%. It could possibly be less, if the conditional probability that given breast cancer, the mammogram being positive is not 100% (i.e. false negatives). This shows the value of correctly understanding and applying Bayes' mathematical theorem.

17.7 SUMMARY:

In this section we proved Bayes theorem and Bayes theorem for future events .Discussed its significance and uses. Examples are discussed to throw light how Bayes theorem can be used to find posterior probability.

17.8 Self Assessment:

(1)In 1989 there were three candidates for the position of principal-Mr. Chatterji,Mr. Ayanger and Dr. singh –whose chance of getting appointment are in the proportion 4:2:3 respectively. The probability that Mr. Chatterji if selected would introduce co-education in the college is 0.3. The probability of Mr. Ayanger and Dr. singh doing the same are respectively 0.5 and 0.8 .What is the probability that there will be co-education in the college in 1990.

(2)The chance that doctor A will diagnose a disease X correctly is 60%.The chance that a patient will die by his treatment after correct diagnosis is 40% and the chance of death by wrong diagnosis is 70%.A patient of doctor A, who had disease X,died.What is the chance that his disease was diagnosed correctly.

(3) Five men out of 100 and 25 women out of 10,000 are colour blind. A colour blind person is chosen at random. What is the probability of his being male. Assume that the male and female are in equal numbers.

17.9 Further Readings:

1. Theory of Probability, P.Mukhopadhyay
2. An Introduction to Theory of Probability and Mathematical Statistics : V K Rohatgi
3. An Introduction to Probability Theory and its Application,w.Feller
4. Fundamentals of Mathematical Statistics : S C Gupta and V K Kapoor
5. New Mathematical Statistics : Bansi Lal and Sanjay Arora

****

UNIT-V **LESSON-18**

PROBABILITY DENSITY FUNCTION

18.1. Introduction

In application of probabilities, we are often concerned with numerical values which are random in nature. For example, we may consider the number of customers arriving at service station at a particular interval of time or the transmission time of a message in a communication system. These random quantities may be considered as real-valued function on the sample space. Such a real-valued function is called real random variable and plays an important role in describing random data. We shall introduce the concept of random variables and related density functions in this lesson.

18.2 Objective

1.To introduce the concept of random variable

2. To introduce Probability Density Function

3.To introduce the properties of Probability density functions

18.3 Random Variable

Mathematical Preliminaries

Real-valued point function on a set

Recall that a real-valued function $f: S \to R$ maps each element $s \in S$, a unique element $f(s) \in$ R .The set S is called the *domain* of f and the set $R_f = \{f(x) \mid x \in S\}$ is called the *range* of f. Clearly $R_f \subseteq$ R. The range and domain of f are shown in Fig,.

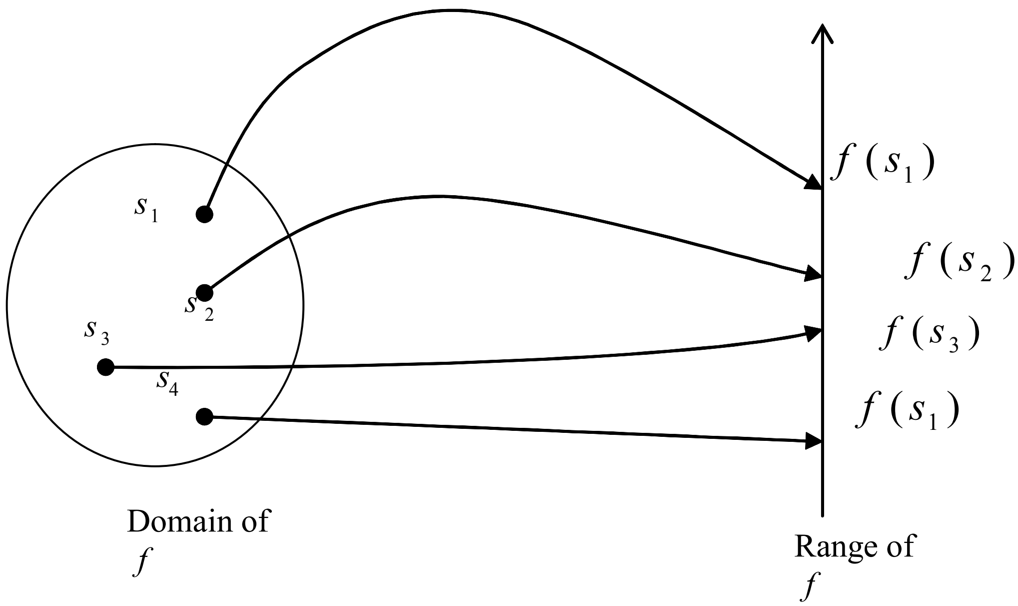


Image and Inverse image

For a point $s \in S$, the functional value *f(s)* $\in$ R. is called the *image* of the point s. If $A \subseteq S$, then the set of the images of elements of A is called *the image of* A and denoted by $f(A)$. Thus

$f(A) = \{f(s) \mid s \in A\}$

Clearly $f(A) \subseteq R_f$

Suppose B $\subseteq$ R_1 The set $\{x \mid f(x) \in \mathrm{B}\}$ is called the inverse image of B under f and is denoted by $f^{-1}(\mathrm{B})$.

Example 5.3.1. Suppose $S = \{H, T\}$ and f : S $\rightarrow$ R is defined by $f(H) = 1$ and $f(T) = -1$. Therefore,

- Rf = {1, -1) $\subseteq$ R
- Image of H is 1 and that of T is -1.
- For a subset of R say $\mathrm{B}_1 = (-\infty,\ 1.5]$,

$f^{-1}(\mathrm{B}_1) = \{s \mid f(s) \in \mathrm{B}_1\} = \{H, T\}$.

For another subset $B_2 = [5,\ 9], f^{-1}(B_2) = \phi$.

Continuous function

A real function $f : \mathrm{S} \to \mathrm{R}$ is said to be continuous at a point a if and only if

(i) $f(a)$ is defined

(ii) $\lim_{x \to a+} f(x) = \lim_{x \to a-} f(x) = f(a)$

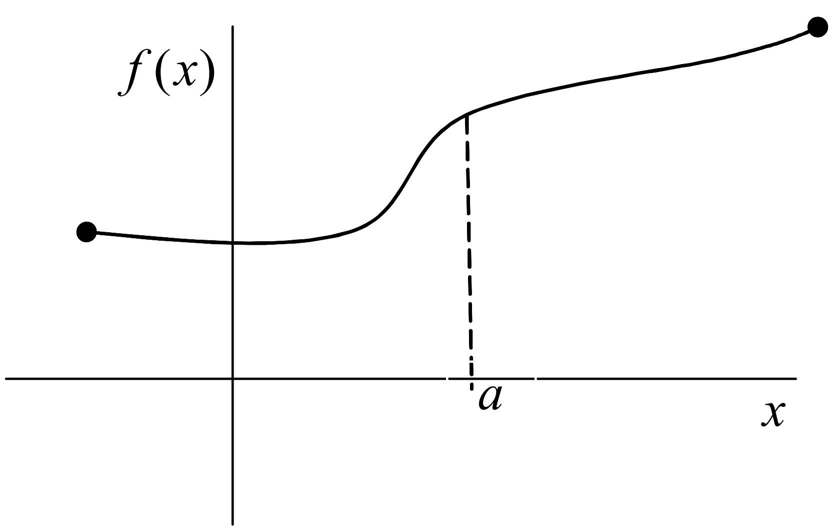
The function $f : \mathrm{S} \to \mathrm{R}$ is said to be right-continuous

at a point a if and only if

(iii) $f(a)$ is defined

$\lim_{x \to a+} f(x) = f(a)$

We can similarly define a left-continuous function.



Function continuous at $x = a$

18.3.1 Random variable

A random variable associates the points in the sample space with real numbers.

Consider the probability space (S, F, P) and function $X: \mathrm{S} \rightarrow \mathrm{R}$ mapping the sample space S into the real line. Let us define the probability of a subset $\mathrm{B} \subseteq \mathrm{S}$ by

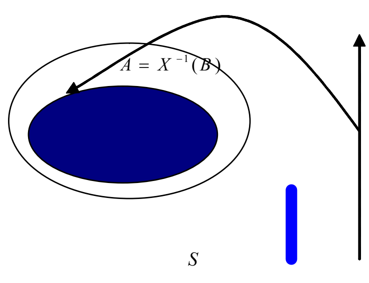
$P_X(\{B\}) = P(X^{-1}(B)) = P(\{s \mid X(s) \in B\})$. Such a definition will be valid if $(X^{-1}(B))$ is a valid event. If S is a discrete sample space, $(X^{-1}(B))$ is always a valid event, but the same may not be true if S is infinite. The concept of sigma algebra is again necessary to overcome this difficulty. We also need the Borel sigma algebra B -the sigma algebra defined on the real line.

The function $X: \mathrm{S} \rightarrow \mathrm{R}$ called a *random variable* if the inverse image of all Borel sets under X is an event. Thus, if X is a random variable, then

$X^{-1}(\mathrm{B}) = \{s \mid X(s) \in \mathrm{B}\} \in \mathrm{F}$.

Remark

- S is the domain of X.
- The *range* of X, denoted by R_X, is given by



Notations:

- Random variables are represented by upper-case letters.
- Values of a random variable are denoted by lower case letters
- $X(s) = x$ means that x is the value of a random variable X at the sample point s.

s

$X = x$.

Figure

Remark

- S is the domain of X.
- The *range* of X, denoted by R_X, is given by

$$R_X = \{X(s) \mid s \in S\}.$$

Clearly $R_X \subseteq$ R.

- The above definition of the random variable requires that the mapping X is such that $X^{-1}(\mathrm{B})$ is a valid event in S. If S is a discrete sample space, this requirement is met by any mapping X: S $\rightarrow$ R. Thus *any mapping defined on the discrete sample space is a random variable.*

Example 5.3.2: Consider the example of tossing a fair coin twice. The psample space is S={ HH,HT,TH,TT} and all four outcomes are equally likely. Then we can define a random variable X as follows

Sample Point	Value of the random variable $X = x$
HH	0
HT	1
TH	2
TT	3

Here $R_X = \{0,1,2,3\}$.

Example 5.3.3.: Consider the sample space associated with the single toss of a fair die. The sample space is given by $S = \{1,2,3,4,5,6\}$

If we define the random variable X that associates a real number equal to the number in the face of the die, then

$X = \{1,2,3,4,5,6\}$

18.4. Probability Space induced by a Random Variable

The random variable X induces a probability measure P_X on $\mathcal{B}$ defined by

$$P_X(\{B\}) = P(X^{-1}(B)) = P(\{s \mid X(s) \in B\})$$

The probability measure P_X satisfies the three axioms of probability:

Axiom 1

$$P_X(B) = P(X^{-1}(B)) \le 1$$

Axiom 2

$$P_X(S) = P(X^{-1}(S)) = P(S) = 1$$

Axiom 3

Suppose $B_1, B_2, ...$ are disjoint Borel sets. Then $X^{-1}(B_1), X^{-1}(B_2), ...$ are distinct events in $\mathcal{F}$. Therefore,

$$\begin{aligned} P_X(\bigcup_{i=1}^{\infty} B_i) &= P(\bigcup_{i=1}^{\infty} X^{-1}(B_i)) \\ &= \sum_{i=1}^{\infty} P(X^{-1}(B_i)) \\ &= \sum_{i=1}^{\infty} P_X(B_i) \end{aligned}$$

Thus the random variable X induces a probability space $(S, \mathcal{B}, P_X)$

18.5 Probability Distribution Function

We have seen that the event B and $\{s \mid X(s) \in B\}$ are equivalent and $P_X(\{B\}) = P(\{s \mid X(s) \in B\})$. The underlying sample space is omitted in notation and we simply write $\{X \in B\}$ and $P(\{X \in B\})$ in stead of $\{s \mid X(s) \in B\}$ and $P(\{s \mid X(s) \in B\})$ respectively.

Consider the Borel set $(-\infty, x]$ where x represents any real number. The equivalent event $X^{-1}((-\infty, x]) = \{s \mid X(s) \leq x, \ \mathrm{s} \in \mathrm{S}\}$ is denoted as $\{X \leq x\}$ The event $\{X \leq x\}$ can be taken as a representative event in studying the probability description of a random variable X. Any other event can be represented in terms of this event. For example, and so on.

$$\{X > x\} = \{X \leq x\}^c, \{x_1 < X \leq x_2\} = \{X \leq x_2\} \setminus \{X \leq x_1\},$$
$$\{X = x\} = \bigcap_{n=1}^{\infty} \left(\{X \leq x\} \setminus \{X \leq x - \frac{1}{n}\} \right)$$

The probability $P(\{X \leq x\}) = P(\{s \mid X(s) \leq x, \ s \in S\})$ is called the *probability distribution function* (also called the *cumulative distribution function* abbreviated as *CDF*) of X and denoted by $F_X(x)$. Thus $(-\infty, x]$,

$$F_X(x) = P(\{X \leq x\})$$

Eaxmple 18.5.1. Consider the random variable X in Example 1

We have

Value Value of the random variable $X = x$	$P(\{X = x\})$
0	$\frac{1}{4}$
1	$\frac{1}{4}$
2	$\frac{1}{4}$
3	$\frac{1}{4}$

For $x < 0$,

$$F_X(x) = P(\{X \le x\}) = 0$$

For $0 \le x < 1$,

$$F_X(x) = P(\{X \le x\}) = P(\{X = 0\}) = \frac{1}{4}$$

For $1 \le x < 2$,

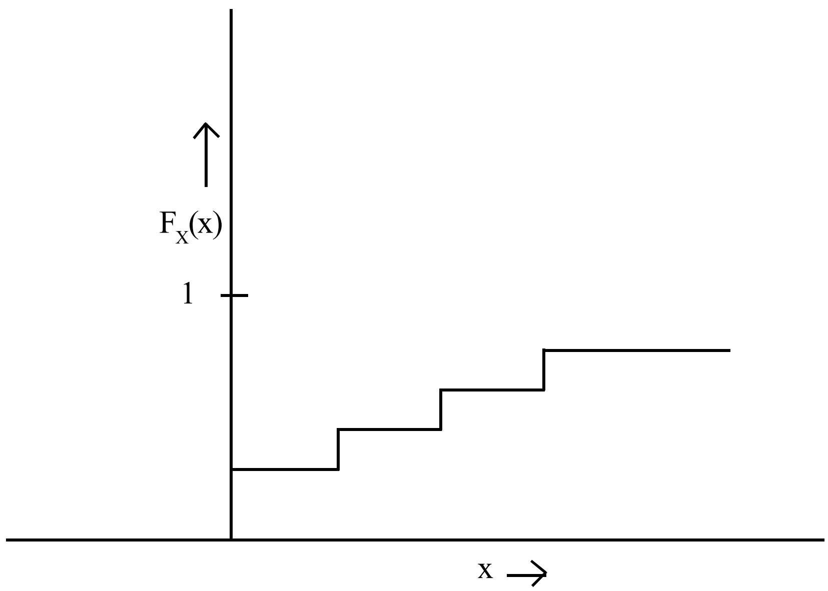
$$\begin{aligned} F_X(x) &= P(\{X \le x\}) \\ &= P(\{X = 0\} \cup \{X = 1\}) \\ &= P(\{X = 0\}) + P(\{X = 1\}) \\ &= \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \end{aligned}$$

For $2 \le x < 3$,

$$F_X(x) = P(\{X \le x\})$$
$$= P(\{X = 0\} \cup \{X = 1\} \cup \{X = 2\})$$
$$= P(\{X = 0\}) + P(\{X = 1\}) + P(\{X = 2\})$$
$$= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$

For $x \ge 3$,

$$F_X(x) = P(\{X \le x\})$$
$$= P(S)$$
$$= 1$$



18.2.Properties of Distribution Function

- $0 \le F_X(x) \le 1$

This follows from the fact that $F_X(x)$ is a probability and its value should lie between 0 and 1.

- $F_X(x)$ is a *non-decreasing function* of X. Thus, $x_1 < x_2$, then $F_X(x_1) < F_X(x_2)$

$x_1 < x_2$

$\Rightarrow \{X(s) \le x_1\} \subseteq \{X(s) \le x_2\}$

$\Rightarrow P\{X(s) \le x_1\} \le P\{X(s) \le x_2\}$

$\therefore\ F_X(x_1) < F_X(x_2)$

- $F_X(x)$ is right continuous

$$F_X(x^+) = \lim_{\substack{h\to 0\\ h>0}} F_X(x+h) = F_X(x)$$

Because, $\lim_{\substack{h\to 0\\ h>0}} F_X(x+h) = \lim_{\substack{h\to 0\\ h>0}} P\{X(s) \le x+h\}$

$= P\{X(s) \le x\}$

$= F_X(x)$

- $F_X(-\infty) = 0$

Because, $F_X(-\infty) = P\{s \mid X(s) \le -\infty\} = P(\phi) = 0$

- $F_X(\infty) = 1$

Because, $F_X(\infty) = P\{s \mid X(s) \le \infty\} = P(S) = 1$

$$P(\{x_1 < X \le x_2\}) = F_X(x_2) - F_X(x_1)$$

We have

$\{X \le x_2\} = \{X \le x_1\} \cup \{x_1 < X \le x_2\}$

$\therefore P(\{X \le x_2\}) = P(\{X \le x_1\}) + P(\{x_1 < X \le x_2\})$

$\Rightarrow P(\{x_1 < X \le x_2\}) = P(\{X \le x_2\}) - P(\{X \le x_1\}) = F_X(x_2) - F_X(x_1)$

- $F_X(x^-) = F_X(x) - P(X = x)$

$F_X(x^-) = \lim_{\substack{h\to 0\\ h>0}} F_X(x-h)$

$= \lim_{\substack{h\to 0\\ h>0}} P\{X(s) \le x-h\}$

$= P\{X(s) \le x\} - P(X(s) = x)$

$= F_X(x) - P(X = x)$

We can further establish the following results on probability of events on the real line:

$$P\{x_1 \le X \le x_2\} = F_X(x_2) - F_X(x_1) + P(X = x_1)$$

$$P(\{x_1 \le X < x_2\}) = F_X(x_2) - F_X(x_1) + P(X = x_1) - P(X = x_2)$$

$$P(\{X > x\}) = P(\{x < X < \infty\}) = 1 - F_X(x)$$

Thus we have seen that given $F_X(x),\ -\infty < x < \infty,$ we can determine the probability of any event involving values of the random variable X. Thus $F_X(x)\ \forall \mathrm{x} \in \mathrm{X}$ is a complete description of the random variable X.

Example 18.5.2. Consider the random variable X defined by

$$\begin{aligned} F_X(x) &= 0, & x &< -2 \\ &= \frac{1}{8}x + \frac{1}{4} & -2 &\le x < 0 \\ &= 1, & x &\ge 0 \end{aligned}$$

Find a) $P(X = 0)$

b) $P\{X \le 0\}$

c) $P\{X > 2\}$

d) $P\{-1 < X \le 1\}$

Solution:

a) $P(X = 0) = F_X(0^+) - F_X(0^-)$

$$= 1 - \frac{1}{4} = \frac{3}{4}$$

b) $P\{X \le 0\} = F_X(0)$

$$= 1$$

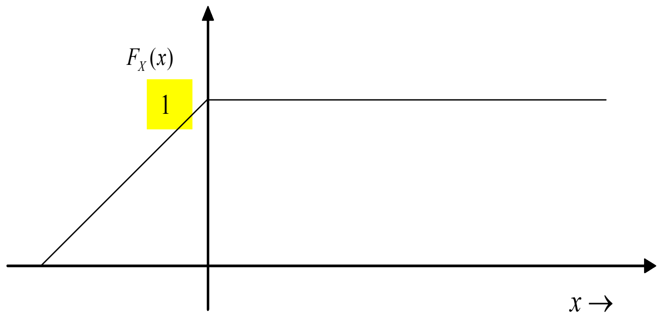
c) $P\{X > 2\} = 1 - F_X(2)$

$= 1 - 1 = 0$

d) $P\{-1 < X \le 1\}$

$= F_X(1) - F_X(-1)$

$= 1 - \frac{1}{8} = \frac{7}{8}$



Example 18.5.7. A coin weighted so that {P(H) = 3/4 and P(T) = 1/4} is tossed three times. Let X be the random variable which denotes the longest string of heads which occurs.Find the distribution of X.

Solution:The random variable X is defined on the sample space

S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

The points in S have the following respective probabilities:

P(HHH) = $(3/4)^3$=27/64

P(HHT = $(3/4)^2$*(1/4)=9/64

P(HTH) = (3/4)*(1/4)*(3/4)=9/64

P(HTT) = (3/4)*$(1/4)^2$=3/64

P(THH) = (1/4)*$(3/4)^2$=9/64

P(THT) = 3/64

P(TTH)=3/64

P(TTT)=1/64

Since X denotes the longest string of heads,

X(TTT) = 0; X(HTT) = 1, X(HTH) = 1, X(THT) = 1, X(TTH) = 1;

X(HHT) = 2, X(THH) = 2; X(HHH) = 3

Thus the image set of X is X(S) = {0, 1, 2, 3}. The probability f (x_i) of each number x_i in X(S) is obtained by summing the probabilities of the points in S whose image is x_i:

f(0) = P(TTT) = 1/64

f(1) = P(HTT) + P(HTH) + P(THT) + P(TTH) = 18/64

f(2) = P(HHT) + P(THH) = 18/64

f(3) = P(HHH)=27/64

5.6. Lesson End Exercises

1. A fair die is tossed. Let X denote twice the number appearing, and let Y denote 1 or 3 according as an odd or an even number appears. Find the distribution (i) X, (ii) Y, (iii) X + Y, (iv) XY.

18.7. Suggested Readings

1. Goon,Gupta;Das Gupta (1991).Fundamental of Statistics
2. Gupta,S.C. and Kapoor,V.K. Fundamental of Mathematical Statistics
3. Hoel,P.G. (1971).Introductory of Mathematical Statistics
4. Hogg R.V and Craig,A.T.Introduction to Mathematical Statistics
5. Hogg ,RV and Tanis,EA(1993).Probability and Statistical Inference
6. Mood,AM,Bose DC and Graybill F A.Introduction to the Theory of Statistics
7. Rohtagi,VK.An Introduction to Probability Theory and Mathematical Statistics

****

UNIT-V **LESSON-19**

DISCRETE & CONTINOUS RANDOM VARIABLES

19.1 Introduction

Random variable is the backbone of the Probability Theory.In the last lesion we have defined the random variable on sample space.In this lesson ,we will classify and define the two types of random variable viz. Discrete and Continuous.Also we will study the related density functions with examples.

19.2.Objectives

1. To introduce Discrete and Continous random variables
2. To introduce probability mass function
3. To introduce Probability density function

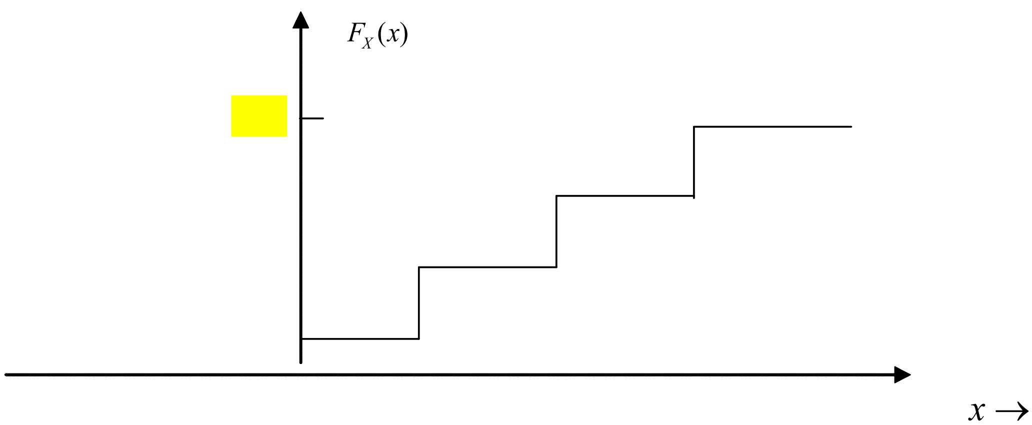
19.3 Discrete, Continuous and Mixed-type random variables

- A random variable X is called a discrete random variable $F_X(x)$ is piece-wise constant. Thus $F_X(x)$ is flat except at the points of jump discontinuity. If the sample space S is discrete the random variable X defined on it is always discrete.
- X is called a *continuous random variable* if $F_X(x)$ is an absolutely continuous function of x. Thus $F_X(x)$ is continuous everywhere on R and $F_X{}'(x)$ exists everywhere except at finite or countably infinite points.
- X is called a *mixed random variable* if $F_X(x)$ has jump

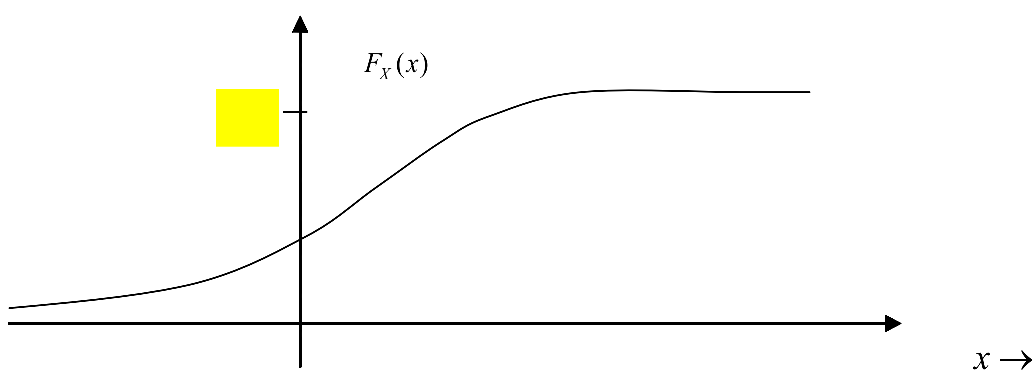
discontinuity at countable number of points and it increases continuously at least at one interval of values of x. For a such type RV X,

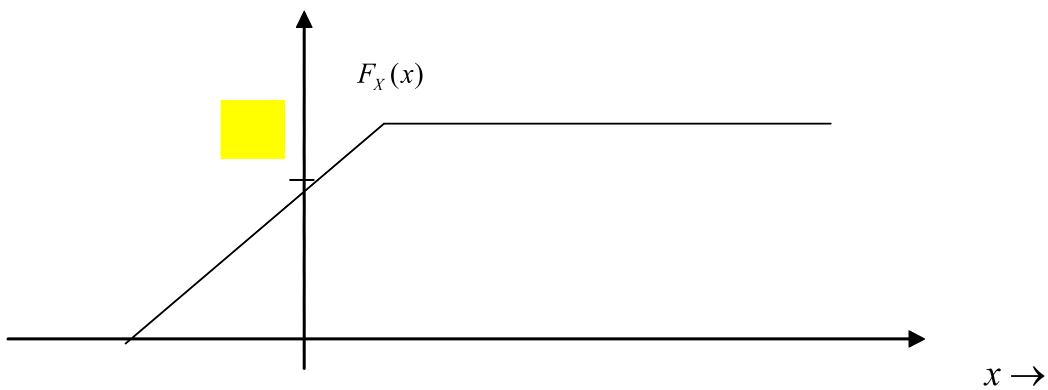
$$F_X(x) = pF_{X_d}(x) + (1-p)F_{X_c}(x)$$

where $F_{X_d}(x)$ is the distribution function of a discrete RV and $F_{X_c}(x)$ is the distribution function of a continuous RV. Typical plots of $F_X(x)$ for discrete, continuous and mixed-random variables are shown in Fig below.

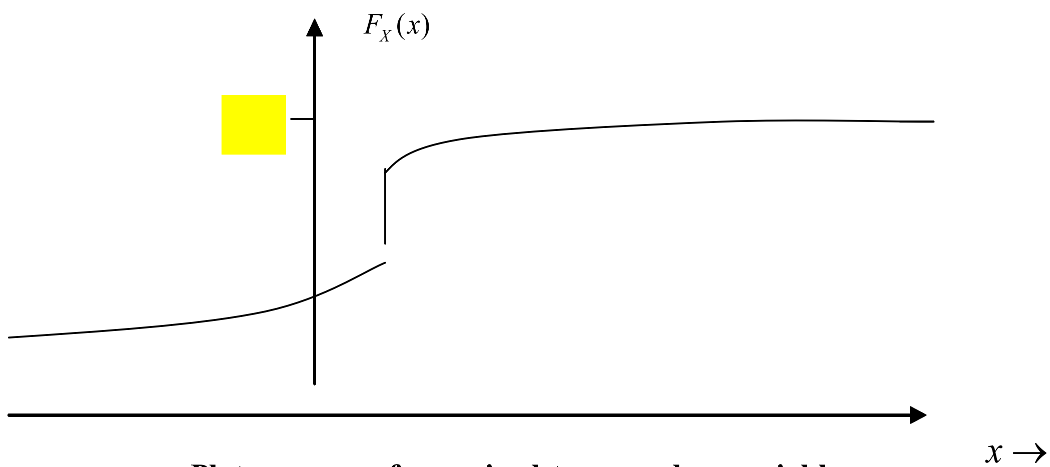


Plot $F_X(x)$ vs. x for a discrete random variable





Plot $F_X(x)$ vs. x **for a continuous random variable**



Plot $F_X(x)$ vs. x **for a mixed-type random variable**

19.4 . Discrete Random Variables and Probability mass functions

A random variable is said to be *discrete* if the number of elements in the range of R_X is finite or countably infinite.

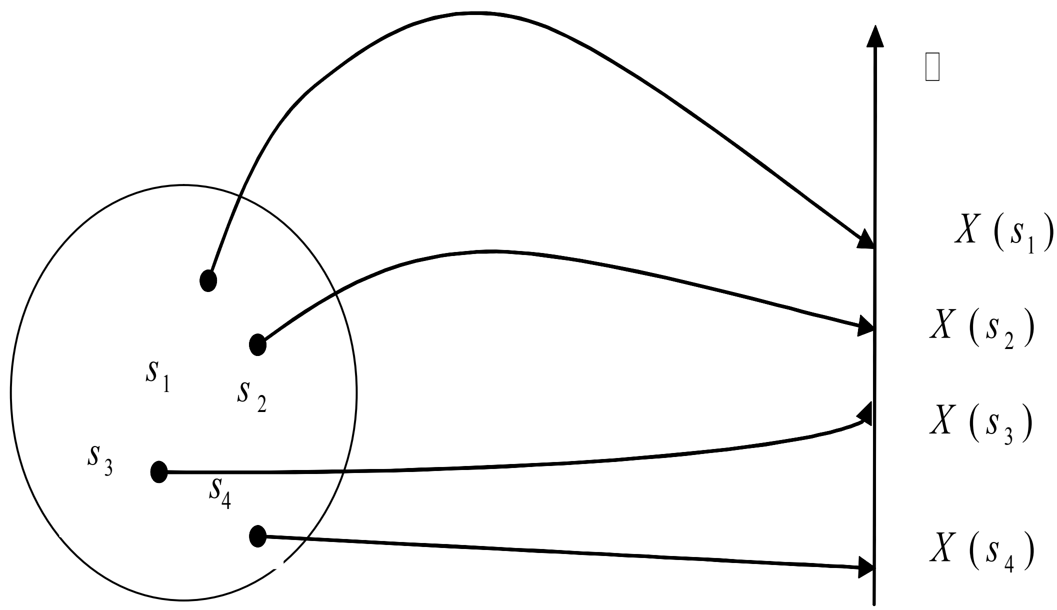
Assume R_X to be countably finite. Let $x_1, x_2, x_3 ..., x_N$ be the elements in R_X. Here the mapping $X(s)$ partitions S into N subsets $\{s \mid X(s) = x_i\}, i = 1, 2, ..., N.$

The discrete random variable in this case is completely specified by the *probability mass function* (pmf) $p_X(x_i) = P(s \mid X(s) = x_i), i = 1, 2, ..., N.$

Clearly,

- $p_X(x_i) \geq 0 \;\; \forall x_i \in R_X$ and
- $\sum_{i \in R_X} p_X(x_i) = 1$
- Suppose $D \subseteq R_X$. Then

$$P(\{x \in D\}) = \sum_{x_i \in D} p_X(x_i)$$

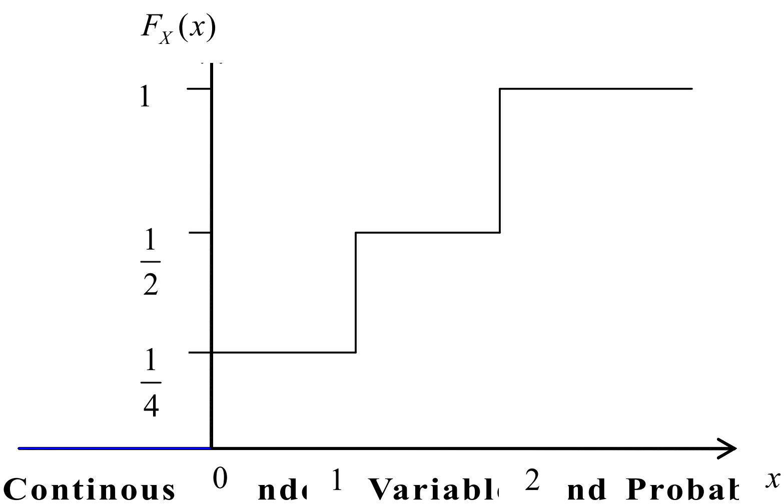


$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{4} & 0 \leq x < 1 \\ \frac{1}{2} & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

The plot of the $F_X(x)$ is shown in Fig. below

The probability mass function of the random variable is given by

We shall describe about some useful discrete probability mass functions in a later class.



19.5. Continous nd Variabl nd Probab y Density Functions For a continuous random variable X, $F_X(x)$ is continuous everywhere. Therefore,

$F_X(x) = F_X(x^-) \quad \forall x \in x$. This implies that

$$\begin{aligned} p_X(x) &= P(\{X = x\}) \\ &= F_X(x) - F_X(x^-) \\ &= 0 \end{aligned}$$

Therefore, the probability mass function of a continuous RV X is zero for all x. A continuous random variable cannot be characterized by the probability mass function. A continuous random variable has a very important chacterisation in terms of a function called the *probability density function.*

Value of RV X = x	**P_X (x)**
0	**1/4**
1	**1/4**
2	**1/2**

If $F_X(x)$ is differentiable, the probability density function (pdf) of X, denoted by $f_X(x)$, us defined as

$$f_X(x) = \frac{d}{dx} F_X(x)$$

Interpretation of $f_X(x)$

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) \\ &= \lim_{\Delta x \to 0} \frac{F_X(x+\Delta x) - F_X(x)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{P(\{x < X \leq x+\Delta x\})}{\Delta x} \end{aligned}$$

so that

$P(\{x < X \leq x + \Delta x\}) = fx(x)\Delta x$

Thus the probability of X lying in the some interval $(x, x+\Delta x]$ is determined by $f_X(x)$. In that sense, $f_X(x)$ represents the concentration of probability just as the density represents the concentration of mass.

Properties of the Probability Density Function

- $f_X(x) \geq 0.$

This follows from the fact that $F_X(x)$ is a non-decreasing function

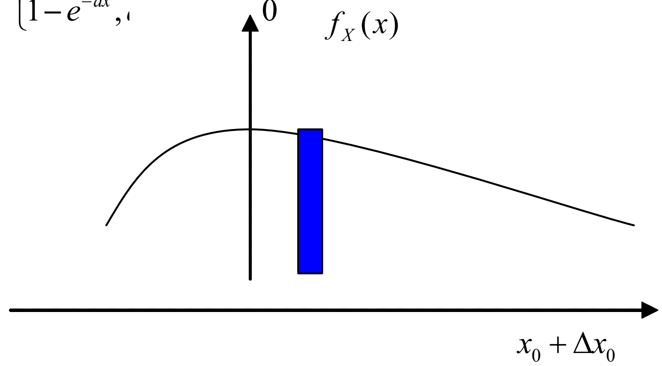
- $F_X(x) = \int_{-\infty}^{x} f_X(u)du$

- $\int_{-\infty}^{\infty} f_X(x)dx = 1$

- $P(x_1 < X \leq x_2) = \int_{-x_1}^{x_2} f_X(x)dx$

Example 19.5.1. Consider the random variable X with the distribution function.

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-ax}, \ldots & 0 \end{cases}$$



The pdf of the RV is given bym

$$f_X(x) = \begin{cases} 0 & x < 0 \\ e^{-ax}, a > 0 & x \geq 0 \end{cases}$$

Remark: Using the Dirac delta function we can define the density function for a discrete random variables.

Consider the random variable X defined by the *probability mass function* (pmf) $p_X(x_i) = P(s \mid X(s) = x_i), i = 1, 2, ..., N.$

The distribution function $F_X(x)$ can be written as

$$F_X(x) = \sum_{i=1}^{N} p_X(x_i) u(x - x_i)$$

where $u(x - x_i)$ shifted unit-step function given by

$$u(x - x_i) = \begin{cases} 1 & \text{for} \quad x \geq x_i \\ 0 & \text{otherwise} \end{cases}$$

Then the density function $f_X(x)$ can be written in terms of the Dirac delta function as

$$f_X(x) = \sum_{i=1}^{n} p_X(x_i) \delta(x - x_i)$$

Example 19.5.2.

Consider the random variable defined in above example. The distribution function $F_X(x)$ can be written as

$$F_X(x) = \frac{1}{4}u(x) + \frac{1}{4}u(x-1) + \frac{1}{2}u(x-2)$$

and

$$f_X(x) = \frac{1}{4}\delta(x) + \frac{1}{4}\delta(x-1) + \frac{1}{2}\delta(x-2)$$

19.6. Probability Density Function of Mixed-type Random Variable

Suppose X is a mixed-type random variable with $F_X(x)$ having jump discontinuity at $X = x_i, i = 1, 2, .., n.$ As already stated, the CDF of a mixed-type random variable X is given by

$$F_X(x) = pF_D(x) + (1-p)F_C(x)$$

where $F_D(x)$ is the conditional distribution function of X given X is discrete and $F_C(x)$ is the conditional distribution function given that X is continuous.

The corresponding pdf is given by

$$f_X(x) = pf_D(x) + (1-p) f_C(x)$$

where

$$f_D(x) = \sum_{i=1}^{n} p_X(x_i)\delta(x - x_i)$$

and $f_C(x)$ is a continuous pdf.

Suppose $R_D = \{x_1, x_2, ..., x_n\}$ denotes the countable subset of points on R_X such that the RV X is characterized by the probability mass function $p_X(x), x \in S_D$. Similarly let $R_C = R_X \setminus R_D$ be a continuous subset of points on R_X such that RV is characterized by the probability density function $f_C(x), x \in R_C$.

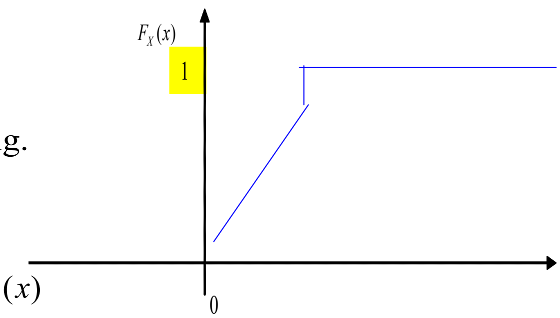
Clearly the subsets R_D and R_C partition the set R_X. If $P(R_D) = p$, then $P(R_C) = 1 - p$.

Thus the probability of the event $\{X \le x\}$ can be expressed as

$$\begin{aligned} P\{X \le x\} &= P(R_D)P(\{X \le x\} \mid R_D) + P(R_C)P(\{X \le x\} \mid R_C) \\ &= pF_D(x) + (1-p)F_C(x) \\ \therefore F_X(x) &= pF_D(x) + (1-p)F_C(x) \end{aligned}$$

Example 19.6.1 Consider the random variable X with the distribution function

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 0.1 & x = 0 \\ 0.1 + 0.8x & 0 < x < 1 \\ 1 & x > 1 \end{cases}$$



The plot of $F_X(x)$ is shown in Fig.

$F_X(x)$ can be expressed as

$$F_X(x) = 0.2F_{X_d}(x) + 0.8F_{X_c}(x)$$

where

$$F_{X_d}(x) = \begin{cases} 0 & x < 0 \\ 0.5 & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

and

$$F_{X_c}(x) = \begin{cases} 0 & x < 0 \\ \text{x} & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

The pdf is given by

$$f_X(x) = 0.2f_{X_d}(x) + 0.8f_{X_c}(x)$$

where

$$f_{X_d}(x) = 0.5\delta(x) + 0.5\delta(x-1)$$

and $f_{X_c}(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases}$

Example 19.6.2.

X is the RV representing the life time of a device with the PDF $f_X(x)$ for x>0. Define the following random variable

$$\begin{aligned} y &= X \quad \text{if} \quad X \le a \\ &= a \quad \text{if} \quad X > a \end{aligned}$$

$R_D = \{a\}$

$R_C = (0, \ a)$

$$\begin{aligned} p &= P\{y \in D\} \\ &= P\{X > a\} \\ &= 1 - F_X(a) \end{aligned}$$

$\therefore F_X(x) = pF_D(x) + (1-p)F_C(x)$

19.7. Lesson End Exercise

Let X be a continuous random variable with distribution with

f(x)=x+k if 0<x<3 and f(x)=0 elsewhere

(i) Evaluate k. (ii) Find P(1<u><=</u> X <= 2)

19.8 Suggested Readings:

1. Goon,Gupta;Das Gupta (1991).Fundamental of Statistics

2. Gupta,S.C. and Kapoor,V.K. Fundamental of Mathematical Statistics

3. Hoel,P.G. (1971).Introductory of Mathematical Statistics

4. Hogg R.V and Craig,A.T.Introduction to Mathematical Statistics
5. Hogg ,RV and Tanis,EA(1993).Probability and Statistical Inference
6. Mood,AM,Bose DC and Graybill F A.Introduction to the Theory of Statistics
7. Rohtagi,VK.An Introduction to Probability Theory and Mathematical Statistics

******

UNIT-V **LESSON-20**

FUNCTIONS OF RANDOM VARIABLES

20.1 Introduction

In the last lesson we have introduced density function of discrete and continuous random variable.But many a times we require the density function of sum of two or more random variables,density functions of product of two or more random variables.So in this lesson we will study the methods to compute density function of function of random variables.This technique is quite useful in computing joint and marginal density functions.

20.2. Objectives

1. To introduce the concept of function of random variable
2. To describe methods to compute density function of function of random variable
3. To give illustrations of derivations
4. To introduce the marginal and joint density functions

20.3. Functions of Random Variables

Often we have to consider random variables which are functions of other random variables. Let X be a random variable and $g(.)$ is a function. Then $Y = g(X)$ is a random variable. We are interested to find the pdf of Y. For example, suppose X represents the random voltage input to a full-wave rectifier. Then the rectifier output Y is given by $Y = |X|$. We have to find the probability description of the random variable Y. We consider the following cases :

(a) X is a discrete random variable with probability mass function $p_X(x)$

The probability mass function of Y is given by

$$
\begin{aligned}
p_Y(y) &= P(\{Y = y\}) \\
&= P(\{x \mid g(x) = y\}) \\
&= \sum_{\{x \mid g(x)=y\}} P(\{X = x\}) \\
&= \sum_{\{x \mid g(x)=y\}} p_X(x)
\end{aligned}
$$

(b) X is a continuous random variable with probability density function $f_X(x)$ and $y = g(x)$ is one-to-one and monotonically increasing

The probability distribution function of Y is given by

$$
\begin{aligned}
F_Y(y) &= P\{Y \le y\} \\
&= P\{g(X) \le y\} \\
&= P\{X \le g^{-1}(y)\} \\
&= P(\{X \le x\})\Big]_{x=g^{-1}(y)} \\
&= F_X(x)\Big]_{x=g^{-1}(y)}
\end{aligned}
$$

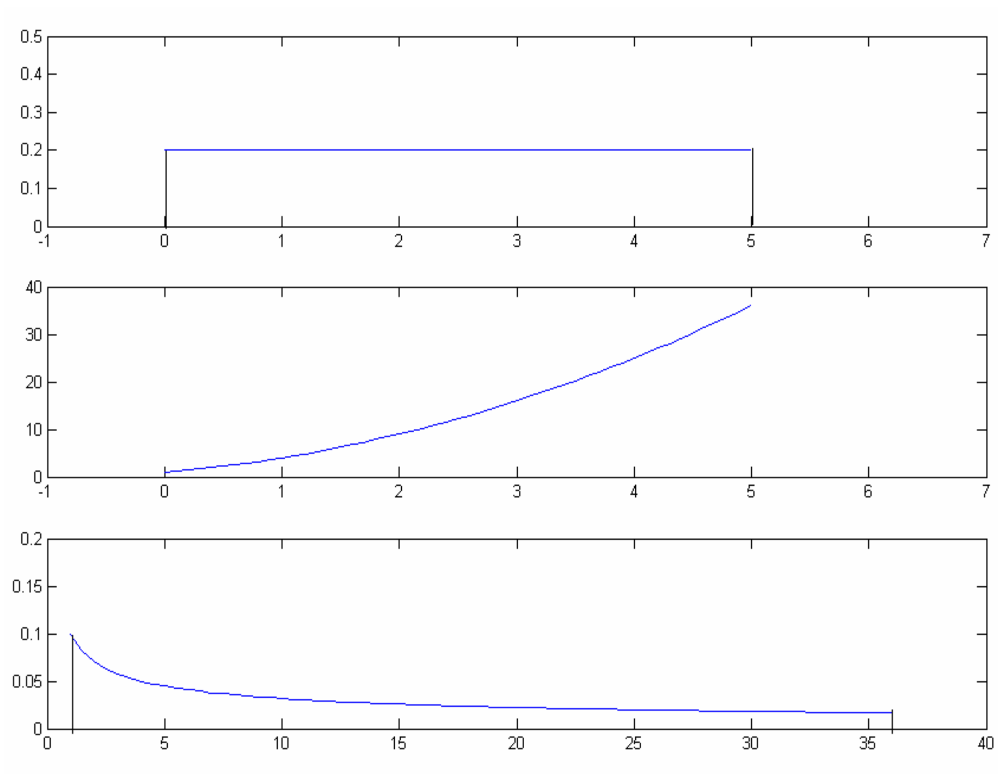
$$
\begin{aligned}
f_Y(y) &= \frac{dF_Y(y)}{dy} \\
&= \frac{dF_X(x)}{dy}\bigg]_{x=g^{-1}(y)} \\
&= \frac{dF_X(x)}{dx}\frac{dx}{dy}\bigg]_{x=g^{-1}(y)}
\end{aligned}
$$

$$= \frac{\frac{dF_X(x)}{dx}}{\frac{dy}{dx}}\Bigg]_{x=g^{-1}(y)}$$

$$= \frac{f_X(x)}{g'(x)}\Bigg]_{x=g^{-1}(y)}$$

$$f_Y(y) = \frac{f_X(x)}{\frac{dy}{dx}} = \frac{f_X(x)}{g'(x)}\Bigg]_{x=g^{-1}(y)}$$

This is illustrated in Fig. below



Example 20.3.1: Probability density function of a linear function of a random variable

Suppose $Y = aX + b,\ a > 0.$

$$\text{Then } x = \frac{y-b}{a} \text{ and } \frac{dy}{dx} = a$$

$$\therefore f_Y(y) = \frac{f_X(x)}{\frac{dy}{dx}} = \frac{f_X(\frac{y-b}{a})}{a}$$

20.4. Probability density function of the distribution function of a random variable

Suppose the distribution function $F_X(x)$ of a continuous random variable X is monotonically increasing and one-to-one and define the random variable $Y = F_X(X)$. Then, $f_Y(y) = 1 \quad 0 \le y \le 1.$

$$y = F_X(x)$$

Clearly $0 \le y \le 1$

$$\frac{dy}{dx} = \frac{dF_X(x)}{dx} = f_X(x)$$

$$\therefore f_Y(y) = \frac{f_X(x)}{\frac{dy}{dx}} = \frac{f_X(x)}{f_X(x)} = 1$$

$$\therefore f_Y(y) = 1 \quad 0 \le y \le 1.$$

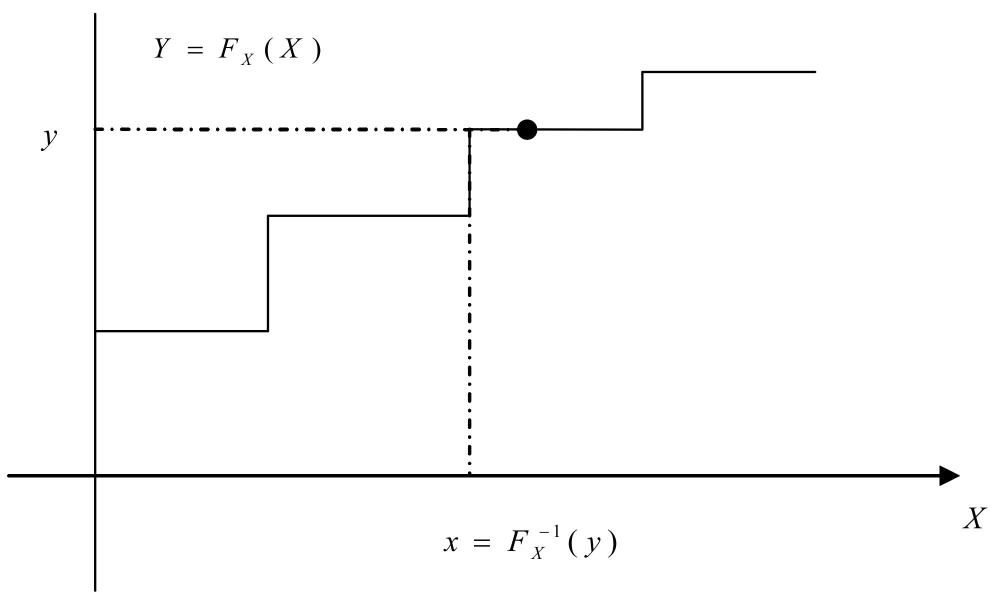
Remark:

(1) The distribution given by $f_Y(y) = 1 \quad 0 \le y \le 1$ is called a uniform distribution over the interval [0,1].

(2) The above result is particularly important in simulating a random variable with a particular distribution function. We assumed $F_X(x)$ to be one-to-one function for invariability. However, the result is more general- *the random variable defined by the distribution function of any random variable is uniformly distributed over* $[0,1]$. For example, if X is a discrete RV,

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P(F_X(x) \le y) \\
&= P(X \le F_X^{-1}(y)) \\
&= F_X(F_X^{-1}(y)) \\
&= y \text{ (Assigning } F_X^{-1}(y) \text{ to the left-most point of the interval for which } F_X(x
\end{aligned}
$$

$$\therefore f_Y(y) = \frac{dF_Y(y)}{dy} = 1 \quad 0 \le y \le 1.$$



(c) X is a continuous random variable with probability density function $f_X(x)$ and $y = g(x)$ has multiple solutions for x

Suppose for $y \in Y, y = g(x)$ has solutions $x_i, i = 1, 2, 3, ..., n$. Then

$$f_Y(y) = \sum_{i=1}^{n} \frac{f_X(x)}{\left|\frac{dy}{dx}\right|_{x=x_i}}$$

Proof:

Consider the plot of $Y = g(X)$. Suppose at a point $y = g(x)$, we have three distinct roots as shown. Consider the event $\{y < Y \leq y + dy\}$. This event will be equivalent to union events

$\{x_1 < X \leq x_1 + dx_1\}, \{x_2 - dx_2 < X \leq x_2\}$ and $\{x_3 < X \leq x_3 + dx_3\}$

$$\therefore P\{y < Y \leq y + dy\} = P\{x_1 < X \leq x_1 + dx_1\} + P\{x_2 - dx_2 < X \leq x_2\} + P\{x_3 < X \leq x_3 + dx_3\}$$

$$\therefore f_Y(y)dy = f_X(x_1)dx_1 + f_X(x_2)(-dx_2) + f_X(x_3)dx_3$$

Where the negative sign in $-dx_2$ is used to account for positive probability.

Therefore, dividing by *dy* and taking the limit, we get

$$\begin{aligned} f_Y(y) &= f_X(x_1)\left(\frac{dx_1}{dy}\right) + f_X(x_2)\left(-\frac{dx_2}{dy}\right) + f_X(x_3)\left(\frac{dx_3}{dy}\right) \\ &= f_X(x_1)\left|\frac{dx_1}{dy}\right| + f_X(x_2)\left|\frac{dx_2}{dy}\right| + f_X(x_3)\left|\frac{dx_3}{dy}\right| \\ &= \sum_{i=1}^{3} \frac{f_X(x_i)}{\left|\frac{dy}{dx}\right|_{x=x_i}} \end{aligned}$$

In the above, we assumed $y = g(x)$ to have three roots. In general, if $y = g(x)$ has *n* roots, then

$$f_Y(y) = \sum_{i=1}^{n} \frac{f_X(x_i)}{\left|\frac{dy}{dx}\right|_{x=x_i}}$$

Example 20.4.1: Probability density function of a linear function of a random variable

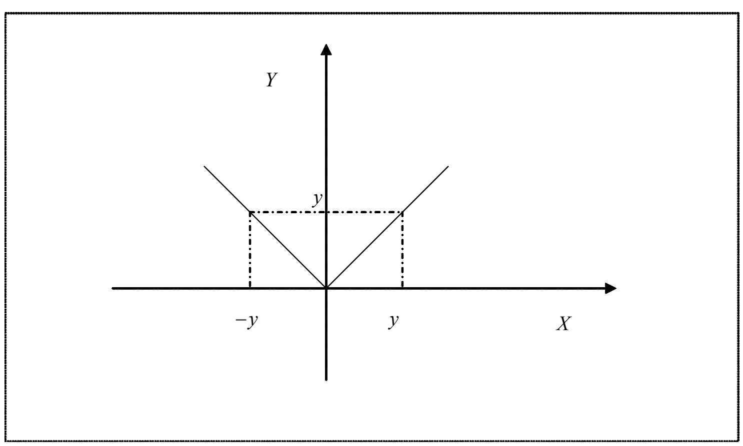
Suppose $Y = aX + b,\ a \neq 0.$

Then $x = \dfrac{y-b}{a}$ and $\dfrac{dy}{dx} = a$

$$\therefore f_Y(y) = \frac{f_X(x)}{\left|\frac{dy}{dx}\right|} = \frac{f_X(\frac{y-b}{a})}{|a|}$$

Example 20.4.2: Probability density function of the output of a full-wave rectifier

Suppose $Y = |X|, \quad -a \leq X \leq a, \qquad a > 0$



$y = |x|$ has two solutions $x_1 = y$ and $x_2 = -y$ and $\left|\frac{dy}{dx}\right| = 1$ at each solution point.

$$\therefore f_Y(y) = \frac{f_X(x)]_{x=y}}{1} + \frac{f_X(x)]_{x=-y}}{1}$$

$$= f_X(y) + f_X(-y)$$

Example 20.4.3.: Probability density function of the output of a square-law device

$Y = cX^2, c \geq 0$

$$\therefore y = cx^2 \qquad => x = \pm\sqrt{\frac{y}{c}} \qquad y \geq 0$$

And $\frac{dy}{dx} = 2cx$ so that $\left|\frac{dy}{dx}\right| = 2c\sqrt{y/c} = 2\sqrt{cy}$

$$\therefore f_Y(y) = \frac{f_X\left(\sqrt{y/c}\right) + f_X\left(\sqrt{-y/c}\right)}{2\sqrt{cy}} \qquad y > 0$$

$= 0$ otherwise

Example 20.4.4: Let X and Y have

$f(x,y) = (3/2)\,(x^2+y^2)$ if $0 < X,Y < 1$

$= 0$ otherwise

Find $E(X^2+Y^2)$.

Solution:

$$E(X^2+Y^2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x^2+y^2) f(x,y)\, dx dy$$

$$= \int_0^1\int_0^1 (3/2)\,(x^2+y^2)^2\, dx dy$$

$= (3/2) \int_0^1 \int_0^1 (x^4+2x^2y^2+y^4)\, dxdy$

$= 14/15$

20.5 Lesson End Exercise

1. Let X, Y be r.v.'s jointly distributed with p.d.f. f given by f(x, y) = $2/c^2$ if 0 = x = y, 0 = y = c and 0 otherwise.

i) Determine the constant c;

ii) Find the marginal p.d.f.'s of X and Y;

iii) Find the conditional p.d.f. of X, given Y, and the conditional p.d.f. of Y, given X;

iv) Calculate the probability that X = 1.

2. Let the r.v.'s X, Y be jointly distributed with p.d.f. f given by f(x, y) = $e^{-x-y} I_{(0,8)\times(0,8)}$ (x, y). Compute the following probabilities:

i) P(X = x);

ii) P(Y = y);

iii) P(X < Y);

iv) P(X + Y = 3).

Exercise.3 Example 8.4: With the following joint pdf

$f(x,y) = kxy^2 \qquad 0 \le x \le y \le 1$

(a) What is k?

(b) Find the marginal pdf.

20.6. Suggested Readings

1. Goon,Gupta;Das Gupta (1991).Fundamental of Statistics
2. Gupta,S.C. and Kapoor,V.K. Fundamental of Mathematical Statistics
3. Hoel,P.G. (1971).Introductory of Mathematical Statistics
4. Hogg R.V and Craig,A.T.Introduction to Mathematical Statistics
5. Hogg ,RV and Tanis,EA(1993).Probability and Statistical Inference
6. Mood,AM,Bose DC and Graybill F A.Introduction to the Theory of Statistics
7. Rohtagi,VK.An Introduction to Probability Theory and Mathematical Statistics

****

UNIT-V **LESSON-21**

MATHEMATICAL EXPECTATIONS

21.1 Introduction

- The *expectation* operation extracts a few parameters of a random variable and provides a summary description of the random variable in terms of these parameters.
- It is far easier to estimate these parameters from data than to estimate the distribution or density function of the random variable.

21.2 Objectives

1. To introduce the concept of Expectation of a random variable
2. To prove properties of expectation of a random variable
3. To compute the expectation of a given random variable having some density function or for given values

21.3 Expected value or mean of a random variable

The expected value $E(x)$ of a random variable X is defined by

$$EX = \int_{-\infty}^{\infty} x f_X(x) dx$$

provided $\int_{-\infty}^{\infty} x f_X(x) dx$ exists.

E(x) is also called the mean or statistical average of the random variable X and denoted by μ_X.

Note that, for a discrete RV X defined by the *probability mass function* (pmf)

$p_X(x_i), i = 1, 2, ..., N,$ the pdf $f_X(x)$ is given by

$$f_X(x) = \sum_{i=1}^{N} p_X(x_i)\delta(x - x_i)$$

$$\begin{aligned}\therefore \mu_X = EX &= \int_{-\infty}^{\infty} x \sum_{i=1}^{N} p_X(x_i)\delta(x - x_i)dx \\ &= \sum_{i=1}^{N} p_X(x_i) \int_{-\infty}^{\infty} x\delta(x - x_i)dx \\ &= \sum_{i=1}^{N} x_i p_X(x_i)\end{aligned}$$

Thus for a discrete random variable X with $p_X(x_i), i = 1, 2, ..., N,$

$$\mu_X = \sum_{i=1}^{N} x_i p_X(x_i)$$

Interpretation of the mean

- The mean gives an idea about the average value of the random value. The values of the random variable are spread about this value.
- Observe that

$$\begin{aligned}\mu_X &= \int_{-\infty}^{\infty} x f_X(x)dx \\ &= \frac{\int_{-\infty}^{\infty} x f_X(x)dx}{\int_{-\infty}^{\infty} f_X(x)dx} \qquad \because \int_{-\infty}^{\infty} f_X(x)dx = 1\end{aligned}$$

Therefore, the mean can be also interpreted as the *centre of gravity* of the pdf curve.

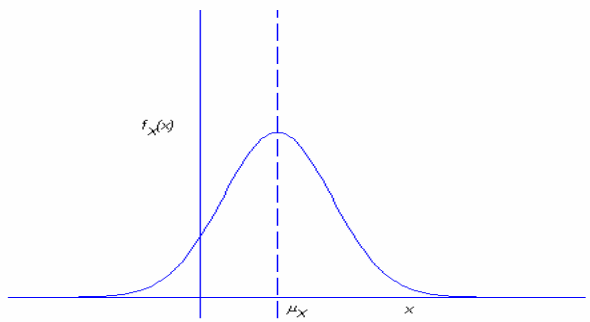


Fig. Mean of a random variable

Example 21.3.1 Suppose X is a random variable defined by the pdf

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\begin{aligned} EX &= \int_{-\infty}^{\infty} x f_X(x)dx \\ &= \int_a^b x \frac{1}{b-a} dx \\ &= \frac{a+b}{2} \end{aligned}$$

Value of the random variable x	0	1	2	3
$p_X(x)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$

Example 21.3.2 Consider the random variable X with pmf as tabulated below

$$\therefore \mu_X = \sum_{i=1}^{N} x_i p_X(x_i)$$

$$= 0 \times \frac{1}{8} + 1 \times \frac{1}{8} + 2 \times \frac{1}{4} + 3 \times \frac{1}{2}$$

$$= \frac{17}{8}$$

Remark If $f_X(x)$ is an even function of x, then $\int_{-\infty}^{\infty} x f_X(x)dx = 0.$ Thus the mean of a RV with an even symmetric pdf is 0.

21.4 Expected value of a function of a random variable

Suppose $Y = g(X)$ is a function of a random variable X as discussed in the last class.

Then, $EY = Eg(X) = \int_{-\infty}^{\infty} g(x) f_X(x)dx$
$E(Y)$

We shall illustrate the theorem in the special case $g(X)$ when $y = g(x)$ is one-to-one and monotonically increasing function of x. In this case,

$$f_Y(y) = \left.\frac{f_X(x)}{g'(x)}\right]_{x=g^{-1}(y)}$$

$$EY = \int_{-\infty}^{\infty} y f_Y(y)dy$$

$$= \int_{y_1}^{y2} y \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y)} dy$$

where $y_1 = g(-\infty)$ and $y_2 = g(\infty)$.

Substituting $x = g^{-1}(y)$ so that $y = g(x)$ and $dy = g'(x)dx,$ we get

$$EY = \int_{-\infty}^{\infty} g(x) f_X(x)dx$$

The following important properties of the expectation operation can be immediately derived:

(a) If c is a constant,

E (c) = C

Clearly

E (c) = C $Ec = \int_{-\infty}^{\infty} cf_X(x)dx = c\int_{-\infty}^{\infty} f_X(x)dx = c$

(b) If $g_1(X)$ and $g_2(X)$ are two functions of the random variable X and c_1 and c_2 are constants,

$E[c_1g_1(X) + c_2g_2(X)] = c_1Eg_1(X) + c_2Eg_2(X)$

$$\begin{aligned} E[c_1g_1(X) + c_2g_2(X)] &= \int_{-\infty}^{\infty} c_1[g_1(x) + c_2g_2(x)]f_X(x)dx \\ &= \int_{-\infty}^{\infty} c_1g_1(x)f_X(x)dx + \int_{-\infty}^{\infty} c_2g_2(x)f_X(x)dx \\ &= c_1\int_{-\infty}^{\infty} g_1(x)f_X(x)dx + c_2\int_{-\infty}^{\infty} g_2(x)f_X(x)dx \\ &= c_1Eg_1(X) + c_2Eg_2(X) \end{aligned}$$

The above property means that E is a linear operator.

Mean-square value

E (X^2) $EX^2 = \int_{-\infty}^{\infty} x^2 f_X(x)dx$

Variance

For a random variable X with the pdf $f_X(x)$ and men μ_X, the variance of X is denoted by σ_X^2 and defined as

$$\sigma_X^2 = E(X-\mu_X)^2 = \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)dx$$

Thus for a discrete random variable X with $p_X(x_i), i = 1, 2, ..., N,$

$$\sigma_X^2 = \sum_{i=1}^{N} (x_i - \mu_X)^2 p_X(x_i)$$

The standard deviation of X is defined as

$$\sigma_X = \sqrt{E(X-\mu_X)^2}$$

Example 21.4.3 Find the variance of the random variable discussed in Example 1.

$$\begin{aligned}\sigma_X^2 &= E(X-\mu_X)^2 \\ &= \int_a^b (x - \frac{a+b}{2})^2 \frac{1}{b-a} dx \\ &= \frac{1}{b-a} [\int_a^b x^2 dx - 2 \times \frac{a+b}{2} \int_a^b x dx + \left(\frac{a+b}{2}\right)^2 \int_a^b dx \\ &= \frac{(b-a)^2}{12}\end{aligned}$$

Example 21.4.4 Find the variance of the random variable discussed in Example 2.

As already computed

$$\mu_X = \frac{17}{8}$$

$$\begin{aligned}\sigma_X^2 &= E(X-\mu_X)^2 \\ &= (0-\frac{17}{8})^2 \times \frac{1}{8} + (1-\frac{17}{8})^2 \times \frac{1}{8} + (2-\frac{17}{8})^2 \times \frac{1}{4} + (3-\frac{17}{8})^2 \times \frac{1}{2} \\ &= \frac{117}{128}\end{aligned}$$

Remark

- Variance is a central moment and measure of dispersion of the random variable about the mean.

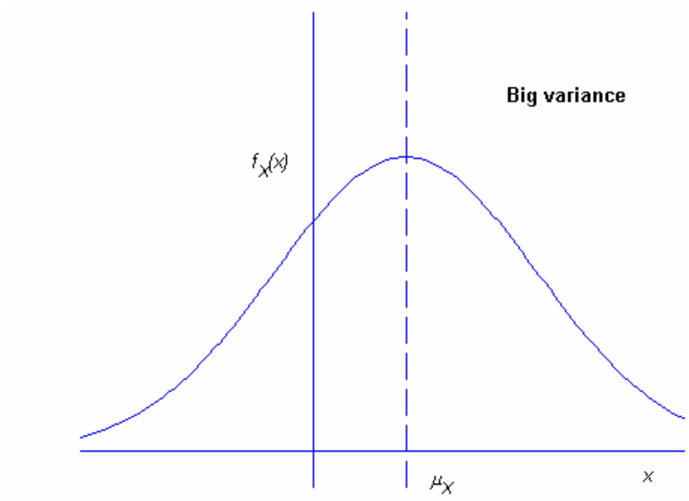
- $E(X-\mu_X)^2$ is the average of the square deviation from the mean. It gives information about the deviation of the values of the RV about the mean. A smaller σ_X^2 implies that the random values are more clustered about the mean, Similarly, a bigger σ_X^2 means that the random values are more scattered.

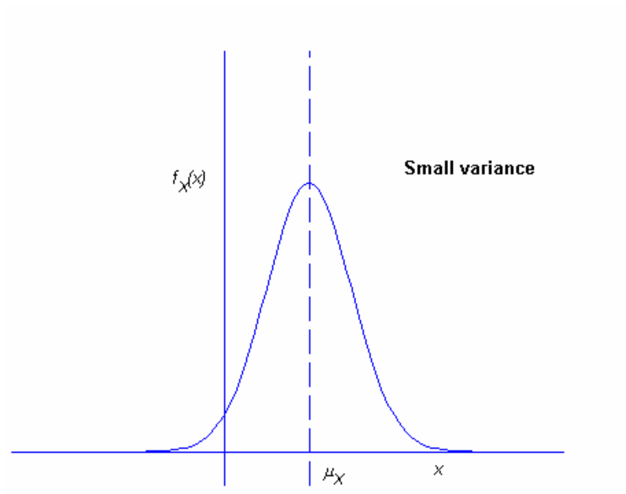
For example, consider two random variables X_1 and X_2 with pmf as shown below. Note that each of X_1 and X_2 has zero means. $\sigma_{X_1}^2=\frac{1}{2}$ and $\sigma_{X_2}^2=\frac{5}{3}$ implying that X_2 has more spread about the mean.

x	-1	0	1
$p_{X_1}(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$

x	-2	-1	0	1	2
$p_{X_2}(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

Fig. shows the pdfs of two continuous random variables with same mean but different variances





- We could have used the *mean absolute deviation* $E|X-\mu_X|$ for the same purpose. But it is more difficult both for analysis and numerical calculations.

Properties of variance

(1) $\sigma_X^2 = EX^2 - \mu_X^2$

$$\begin{aligned}\sigma_X^2 &= E(X - \mu_X)^2 \\ &= E(X^2 - 2\mu_X X + \mu_X^2) \\ &= EX^2 - 2\mu_X EX + E\mu_X^2 \\ &= EX^2 - 2\mu_X^2 + \mu_X^2 \\ &= EX^2 - \mu_X^2\end{aligned}$$

(2) If $Y = cX + b$, where c and b are constants, then $\sigma_Y^2 = c^2\sigma_X^2$

$$\begin{aligned}\sigma_Y^2 &= E(cX + b - c\mu_X - b)^2 \\ &= Ec^2(X - \mu_X)^2 \\ &= c^2\sigma_X^2\end{aligned}$$

$\text{var}(c) = 0.$

21.5 nth moment of a random variable

We can define the *nth* moment and the *nth central*-moment of a random variable X by the following relations

nth-order moment $EX^n = \int_{-\infty}^{\infty} x^n f_X(x)dx \quad n = 1, 2, ..$

nth-order central moment $E(X - \mu_X)^n = \int_{-\infty}^{\infty} (x - \mu_X)^n f_X(x)dx \quad n = 1, 2, ...$

Note that

- The mean $\mu_X = EX$ is the first moment and the mean-square value EX^2 is the second moment
- The first central moment is 0 and the variance $\sigma_X^2 = E(X - \mu_X)^2$ is the second central moment

- The third central moment measures lack of symmetry of the pdf of a random variable. $\frac{E(X-\mu_X)^3}{\sigma_X^3}$ is called the *coefficient of skewness* and If the pdf is symmetric this coefficient will be zero.
- The fourth central moment measures flatness of peakednes of the pdf of a random variable. $\frac{E(X-\mu_X)^4}{\sigma_X^4}$ is called *kurtosis*. If the peak of the pdf is sharper, then the random variable has a higher kurtosis.

21. 6 Lesson End Exercises

1. The continuous random variable ξ is called a uniform random variable if its density function is

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{elsewhere} \end{cases}$$

Show that for this variable, the mean $\mu = \frac{a+b}{2}$ and the variance $\sigma^2 = \frac{(b-a)^2}{12}$.

2.Define moments of the random variable.Show that E(aX+b)=a E(X)+b

21. 7 Suggested Readings:

1. Goon,Gupta;Das Gupta (1991).Fundamental of Statistics
2. Gupta,S.C. and Kapoor,V.K. Fundamental of Mathematical Statistics
3. Hoel,P.G. (1971).Introductory of Mathematical Statistics
4. Hogg R.V and Craig,A.T.Introduction to Mathematical Statistics
5. Hogg ,RV and Tanis,EA(1993).Probability and Statistical Inference
6. Mood,AM,Bose DC and Graybill F A.Introduction to the Theory of Statistics
7. Rohtagi,VK.An Introduction to Probability Theory and Mathematical Statistics

****

UNIT-V **LESSON-22**

JOINT AND MARGINAL PROBABILITY FUNCTIONS

22.1 Introduction

In the last lessons the concept of random variable have been explored and we have discussed the function of random variables .In the last lesson we proved some result regarding the expectation of a random variable.But in general,in statistics we encounter more than one random variable say X and Y.So in this lesson we will learn the concept of joint density function of two random variables and if joint density function is given,we will leran how to compute the marginals.

22.2. Objectives

1. To introduce the concept of joint Probablity density function of random variable
2. To introduce concept of marginals density function of random variable
3. To introduce the techniques of computation of marginal function from given joint probability den $\mathbf{R}^2$ functions.

22.3. Joint Probability Density Function of two functions of two random variables

We consider the transformation $(g_1, g_2) : \mathbf{R}^2 \to \mathbf{R}^2$. We have to find out the joint probability density function $f_{Z_1,Z_2}(z_1, z_2)$ where $Z_1 = g_1(X,Y)$

and $Z_2 = g_2(X,Y)$. We hve to find out the joint probability density function

$f_{Z_1,Z_2}(z_1,z_2)$ where $z_1 = g_1(x,y)$ and $z_2 = g_2(x,y)$. Suppose the inverse mapping relation is

$x = h_1(z_1,z_2)$ and $y = h_2(z_1,z_2)$

Consider a differential region of area $dz_1\, dz_2$ at point (z_1,z_2) in the $Z_1 - Z_2$ plane.

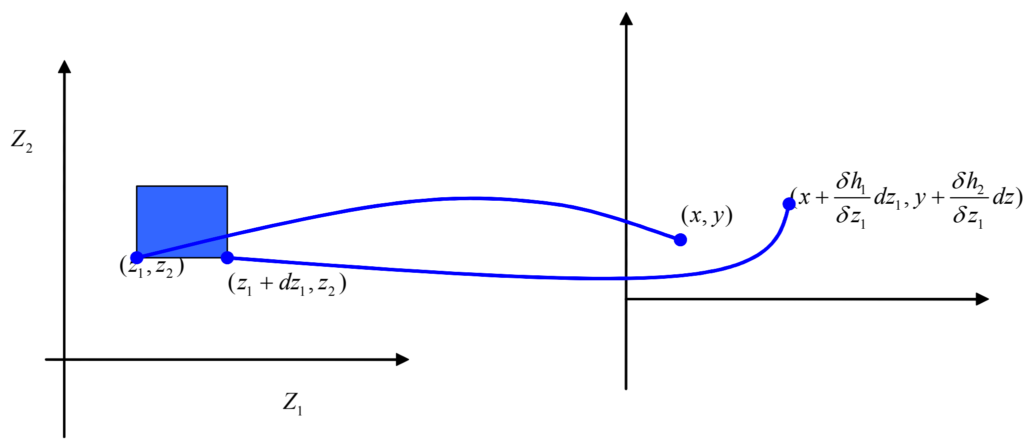
Let us see how the corners of the differential region are mapped to the $X - Y$ plane. Observe that

$$h_1(z_1 + dz_1, z_2) = h_1(z_1,z_2) + \frac{\delta h_1}{\delta z_1} dz_1 = x + \frac{\delta h_1}{\delta z_1} dz_1$$

$$h_2(z_1 + dz_1, z_2) = h_1(z_1,z_2) + \frac{\delta h_2}{\delta z_1} dz_1 = y + \frac{\delta h_2}{\delta z_1} dz_1$$

Therefore,

The point $(z_1 + dz_1, z_2)$ is mapped to the point $(x + \frac{\delta h_1}{\delta z_1} dz_1, y + \frac{\delta h_2}{\delta z_1} dz)$ in the $X - Y$ plane.



We can similarly find the points in the $X-Y$ plane corresponding to (z_1, z_2+dz_2) and (z_1+dz_1, z_2+dz_2). The mapping is shown in Fig. We notice that each differential region in the $X-Y$ plane is a parallelogram. It can be shown that the differential parallelogram at (x, y) has a area $|J(z_1, z_2)|dz_1 dz_2$ where $J(z_1, z_2)$ is the *Jacobian* of the transformation defined as the determinant

$$J(z_1, z_2) = \begin{vmatrix} \frac{\delta h_1}{\delta z_1} & \frac{\delta h_1}{\delta z_2} \\ \frac{\delta h_2}{\delta z_1} & \frac{\delta h_2}{\delta z_2} \end{vmatrix}$$

Further, it can be shown that the absolute values of the Jacobians of the forward and the inverse transform are inverse of each other so that

$$|J(z_1, z_2)| = \frac{1}{|J(x, y)|}$$

where

$$J(x, y) = \begin{vmatrix} \frac{\delta g_1}{\delta x} & \frac{\delta g_1}{\delta y} \\ \frac{\delta g_2}{\delta x} & \frac{\delta g_2}{\delta y} \end{vmatrix}$$

Therefore, the differential parallelogram in Fig. has an area of $\frac{dz_1 dz_2}{|J(x, y)|}$.

Suppose the transformation $z_1 = g_1(x, y)$ and $z_2 = g_2(x, y)$ has n roots and let $(x_i, y_i),\ i = 1, 2, ..n$ be the roots. The inverse mapping of the differential region in the $X-Y$ plane will be n differential regions corresponding to n roots. The inverse mapping is illustrated in the following figure for $n = 4$. As these parallelograms are non- overlapping,

$$f_{Z_1,Z_2}(z_1,z_2)\,dz_1dz_2 = \sum_{i=1}^{n} f_{X,Y}(x,y)\frac{dz_1dz_2}{|J(x_i,y_i)|}$$

$$\therefore \qquad f_{Z_1,Z_2}(z_1,z_2) = \sum_{i=1}^{n} \frac{f_{X,Y}(x,y)}{|J(x_i,y_i)|}$$

Remark

- If $z_1 = g_1(x,y)$ and $z_2 = g_2(x,y)$ does not have a root in (x,y), then $f_{Z_1,Z_2}(z_1,z_2) = 0.$

22.3.2 Jointly distributed random variables

We may define two or more random variables on the same sample space. Let X and Y be two real random variables defined on the same probability space $(S,\mathcal{F},P)$. The mapping $S \to \mathbb{R}^2$ such that for $s \in S,\ (X(s),Y(s)) \in \mathbb{R}^2$ is called a joint random variable.

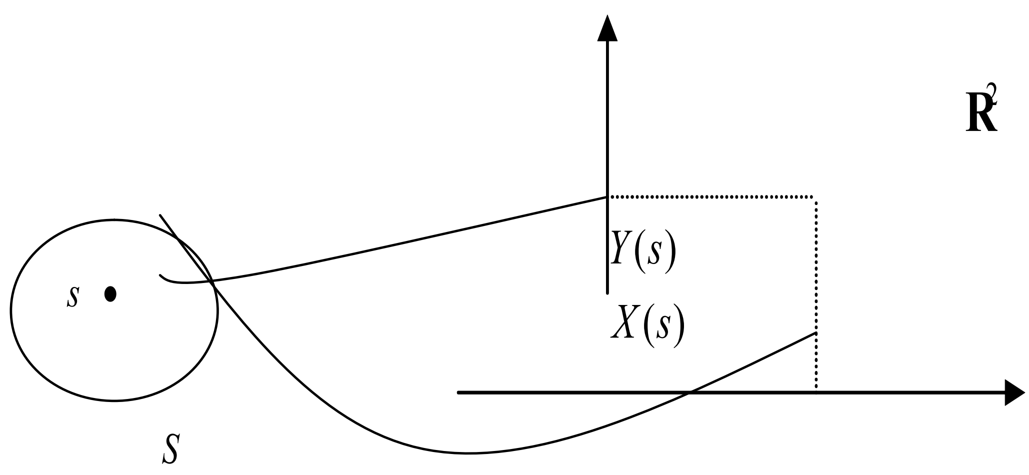


Figure :- Joint Random Variable

Remark

- The above figure illustrates the mapping corresponding to a joint random variable. The joint random variable in the above case is denoted by (X, Y).
- We may represent a joint random variable as a two-dimensional vector $\mathbf{X} = [X\ \mathrm{Y}]'$.
- We can extend the above definition to define joint random variables of any dimension. The mapping $S \to \mathbb{R}^n$ such that for $s \in S,\ (X_1(s), X_2(s), ..., X_n(s)) \in \mathbb{R}^n$ is called a n-dimensional random variable and denoted by the vector $\mathbf{X} = [X_1\ X_2\ ....X_n]'$.

Example 22.3.1: Suppose we are interested in studying the height and weight of the students in a class. We can define the joint RV (X, Y) where X represents height and Y represents the weight.

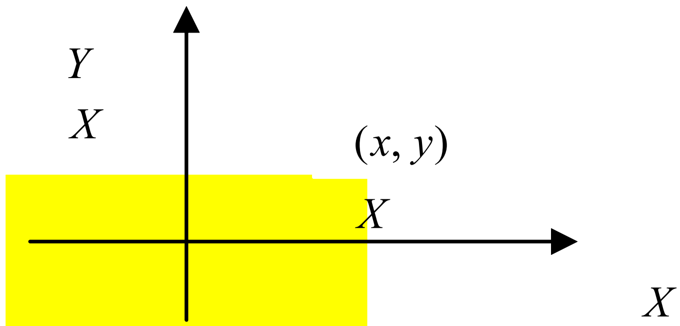
Example 22.3. 2 Suppose in a communication system X is the transmitted signal and Y is the corresponding noisy received signal. Then (X, Y) is a joint random variable.

22.3.3. Joint Probability Distribution Function

Recall the definition of the distribution of a single random variable. The event $\{X \leq x\}$ was used to define the probability distribution function $F_X(x)$. Given $F_X(x)$, we can find the probability of any event involving the random variable. Similarly, for two random variables X and Y, the event $\{X \leq x, Y \leq y\} = \{X \leq x\} \cap \{Y \leq y\}$ is considered as the representative event. The probability $P\{X \leq x, Y \leq y\}\ \forall (x, y) \in \mathrm{R}^2$ is called the *joint distribution function of the random variables* X *and* Y and denoted by $F_{X,Y}(x, y)$.

$F_{X,Y}(x, y)$ satisfies the following properties:

- $F_{X,Y}(x_1, y_1) \leq F_{X,Y}(x_2, y_2)$ if $x_1 \leq x_2$ and $\mathrm{y}_1 \leq y_2$



If $x_1 < x_2$ and $y_1 < y_2$,

$\{X \le x_1, Y \le y_1\} \subseteq \{X \le x_2, Y \le y_2\}$

$\therefore P\{X \le x_1, Y \le y_1\} \le P\{X \le x_2, Y \le y_2\}$

$\therefore F_{X,Y}(x_1, y_1) \le F_{X,Y}(x_2, y_2)$

- $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = 0$

  Note that

  $\{X \le -\infty, Y \le y\} \subseteq \{X \le -\infty\}$
- $F_{X,Y}(\infty, \infty) = 1.$
- $F_{X,Y}(x, y)$ is right continuous in both the variables.
- If $x_1 < x_2$ and $y_1 < y_2$,

Given $F_{X,Y}(x, y), -\infty < x < \infty, -\infty < y < \infty,$ we have a complete description of the random variables X and Y.

- $F_X(x) = F_{XY}(x, +\infty).$

  To prove this

  Similarly $F_Y(y) = F_{XY}(\infty, y).$
- Given $F_{X,Y}(x, y), -\infty < x < \infty, -\infty < y < \infty,$ each of $F_X(x)$ and $F_Y(y)$ is called a marginal distribution function.

Example 22.3.3.

Consider two jointly distributed random variables X and Y with the joint CDF

$$F_{X,Y}(x,y) = \begin{cases} (1-e^{-2x})(1-e^{-y}) & x \geq 0, y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the marginal CDFs

(b) Find the probability $P\{1 < X \leq 2, \ 1 < Y \leq 2\}$

(a) $$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x,y) = \begin{cases} 1-e^{-2x} & x \geq 0 \\ 0 & \text{elsewhere} \end{cases}$$

$$F_Y(y) = \lim_{x \to \infty} F_{X,Y}(x,y) = \begin{cases} 1-e^{-y} & \text{y} \geq 0 \\ 0 & \text{elsewhere} \end{cases}$$

(b)

22.4 Jointly distributed discrete random variables

If X and Y are two discrete random variables defined on the same probability space (S,F,P) such that X takes values from the countable subset R_X and Y takes values from the countable subset R_Y. Then the joint random variable (X,Y) can take values from the countable subset in $R_X \times R_Y$. The joint random variable (X,Y) is completely specified by their *joint probability mass function*

$$p_{X,Y}(x,y) = P\{s \mid X(s) = x, Y(s) = y\}, \quad \forall (x,y) \in R_X \times R_Y.$$

Given $p_{X,Y}(x,y)$, we can determine other probabilities involving the random variables X and Y.

Remark

- $p_{X,Y}(x,y) = 0$ for $(x,y) \notin R_X \times R_Y$

- $\sum_{(x,y)\in R_X\times R_Y} \sum p_{X,Y}(x,y) = 1$

This is because

$$\sum_{(x,y)\in R_X\times R_Y} \sum p_{X,Y}(x,y) = P(\bigcup_{(x,y)\in R_X\times R_Y} \{x,y\})$$
$$=P(R_X \times R_Y)$$
$$=P\{s \mid (X(s),Y(s)) \in (R_X \times R_Y)\}$$
$$=P(S) = 1$$

- *Marginal Probability Mass Functions:* The probability mass functions $p_X(x)$ and $p_Y(y)$ are obtained from the joint probability mass function as follows

$$p_X(x) = P\{X = x\} \cup R_Y$$
$$= \sum_{y\in R_Y} p_{X,Y}(x,y)$$

and similarly

$$p_Y(y) = \sum_{x\in R_X} p_{X,Y}(x,y)$$

These probability mass functions $p_X(x)$ and $p_Y(y)$ obtained from the joint probability mass functions are called *marginal probability mass functions*.

Example 22.4.1. Consider the random variables X and Y with the joint probability mass function as tabulated in Table . The marginal probabilities are as shown in the last column and the last row

22.4.1 Joint Probability Density Function

If X and Y are two continuous random variables and their joint distribution function is continuous in both x and y, then we can define *joint probability density function* $f_{X,Y}(x,y)$ by

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y),$$ provided it exists.

Clearly $F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\,dv\,du$

Properties of Joint Probability Density Function

- $f_{X,Y}(x,y)$ is always a non-negative quantity. That is,

  $f_{X,Y}(x,y) \geq 0 \quad \forall (x,y) \in \mathbb{R}^2$

- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$ **R**

- The probability of any Borel set B can be obtained by

$P(B) = \iint_{(x,y) \in B} f_{X,Y}(x,y)\,dx\,dy$

22.5. Marginal density functions

The marginal density functions $f_X(x)$ and $f_Y(y)$ of two joint RVs X and Y are given by the derivatives of the corresponding marginal distribution functions. Thus

$$\begin{aligned} f_X(x) &= \tfrac{d}{dx} F_X(x) \\ &= \tfrac{d}{dx} F_X(x, \infty) \\ &= \tfrac{d}{dx} \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} f_{X,Y}(u,y)\,dy \right) du \\ &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \end{aligned}$$

and similarly $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$

Remark

- The marginal CDF and pdf are same as the CDF and pdf of the concerned single random variable. The *marginal* term simply refers that it is derived from the corresponding joint distribution or density function of two or more jointly random variables.

- With the help of the two-dimensional Dirac Delta function, we can define the joint pdf of two discrete jointly random variables. Thus for discrete jointly random variables X and Y.

$$f_{X,Y}(x,y) = \sum_{(x_i,y_j)\in R_X\times R_Y.}\sum p_{X,Y}(x,y)\delta(x-x_i, y-y_j)$$

Example 22.5.1. The joint density function $f_{X,Y}(x,y)$ in the previous example is

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{\partial^2}{\partial x\partial y}F_{X,Y}(x,y) \\ &= \frac{\partial^2}{\partial x\partial y}[(1-e^{-2x})(1-e^{-y})] \quad x\ge 0, y\ge 0 \\ &= 2e^{-2x}e^{-y} \quad x\ge 0, y\ge 0 \end{aligned}$$

Example 22.5.2 The joint pdf of two random variables X and Y are given by

$$\begin{aligned} f_{X,Y}(x,y) &= cxy \quad 0\le x\le 2,\ 0\le y\le 2 \\ &= 0 \ \text{ otherwise} \end{aligned}$$

(i) Find c.

(ii) Find $F_{X,y}(x,y)$

(iii) Find $f_X(x)$ and $f_Y(y)$.

(iv) What is the probability $P(0<X\le 1, 0<Y\le 1)$?

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx = c\int_0^2\int_0^2 xydydx$$

$$\therefore c = \frac{1}{4}$$

$$F_{X,Y}(x,y) = \frac{1}{4}\int_0^y\int_0^x uvdudv$$

$$= \frac{x^2y^2}{16}$$

$$f_X(x) = \int_0^2 \frac{xy}{4}dy \quad 0 \le y \le 2$$

$$= \frac{x}{2} \quad 0 \le y \le 2$$

Similarly

$$f_Y(y) = \frac{y}{2} \quad 0 \le y \le 2$$

$$P(0 < X \le 1, 0 < Y \le 1)$$

$$= F_{X,Y}(1,1) + F_{X,Y}(0,0) - F_{X,Y}(0,1) - F_{X,Y}(1,0)$$

$$= \frac{1}{16} + 0 - 0 - 0$$

$$= \frac{1}{16}$$

22.7 Lesson End Exercise

The joint pdf of two random variables X and Y are given by

$$f_{X,Y}(x,y) = cxy \quad 0 \le x \le 2,\ 0 \le y \le 2$$
$$= 0 \quad \text{otherwise}$$

(v) Find c.

(vi) Find $F_{X,y}(x,y)$

(vii) Find $f_X(x)$ and $f_Y(y)$.

What is the probability $P(0<x<1.5, 1<y<2)$?

22.8 Suggested Readings:

1. Goon,Gupta;Das Gupta (1991).Fundamental of Statistics
2. Gupta,S.C. and Kapoor,V.K. Fundamental of Mathematical Statistics
3. Hoel,P.G. (1971).Introductory of Mathematical Statistics
4. Hogg R.V and Craig,A.T.Introduction to Mathematical Statistics
5. Hogg ,RV and Tanis,EA(1993).Probability and Statistical Inference
6. Mood,AM,Bose DC and Graybill F A.Introduction to the Theory of Statistics
7. Rohtagi,VK.An Introduction to Probability Theory and Mathematical Statistics

****

UNIT-V LESSON-23

M.G.F., CONDITIONAL EXPECTATION AND VARIANCE

23.1 Introduction

In this lesson we have introduces the concept of moment generating functions.From mgf of a rv ,we can derive the moments of the distribution.Further we have also discussed the concept of conditional expaectation and conditional variance which play an important role in Regression analysis.Theorems related to condional expaectation and conditional variance have also been proved.

23.2 Objectives:

1. To introduce Moment Generating Function of a rv having density function f(x)
2. To introduce conditional expectation
3. To introduce conditional variance

23.3.Moment generating functions

If X is a r.v., we associate to it a function of a real variable t called the moment generating

function of X , abbreviated m.g.f. This function, which we denote by M_X (t) is defined as

$M_X(t) = E(e^{tX})$.

for those real s for which the expectation exists.(It will always exist if s = 0, and in most cases of interest will exist at least for t in some interval around 0.) Notice that if X has pdf $f_X(x)$ then the moment generating function $M_X(t)$ is defined by

$$M_X(s) = Ee^{SX}$$
$$= \int_{R_X} f_X(x) e^{SX} dx$$

Where R_X is the range of the random variable X.

If X is a non negative continuous random variable, we can write

$$M_X(s) = \int_0^\infty f_X(x) e^{SX} dx$$

The moment generating function $M_X(t)$ will exist only if the sum or integral in the above definition converges. If the moment generating function of a random variable does exist, it can be used to obtain all the origin moments of the random variable.

Remark: 1.The nth moment is equal to the nth derivative of the mgf evaluated at t=0.

i.e. $M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)\Big|_{t=0}$.

2. $$M_x'(s) = \int_0^\infty x f_x(x) e^{sx} dx$$

$$\therefore M'(0) = EX$$

3. $$\frac{d^k}{ds^k} M_X(s) = \int_0^\infty x^k f_X(x) e^{sx} dx$$
$$= E(X^k)$$

Theorem 23.3.1. If X and Y are independent then $M_{X+Y}(t) = M_X(t)M_Y(t)$, that is, the m.g.f. of a sum of independent r.v.'s is the product of their m.g.f's provided these exist. Similarly, the m.g.f of a sum of any finite number of indepedent r.v.'s is the product of their m.g.f's for values of t for which the m.g.f. exists.

Proof. For any fixed t it is plausible (and can be proved) that if X and Y are independent so are e^{tX} and e^{tY} . (In fact, any reasonable function of X and any other reasonable function of Y are independent.) So just using properties of exponents and the fact that the expectation of the product of independent r.v.'s is the product of their expectations we get

$$M_{X+Y}(t) = Ee^{t(X+Y)} = E(e^{tX}e^{tY})$$

$$= E(e^{tX})E(e^{tY})$$

$$= M_X(t)M_Y(t).$$

Example 23.3.1 : Suppose S is binomial r.v. with parameters n and p i.e. S is the number of successes in n independent trials with success probability p on each trial. As we have seen before, we can think of S as the sum of n indicator r.v.'s as in the preceding example:

$$S = X_1 + X_2 + \cdots + X_n .$$

But each of the X_i 's has m.g.f. $q + pe^t$ from the preceding example, so the theorem we just proved says that the m.g.f. of S is the product of these. Thus

$$M_S(t) = (q + pe^t)^n .$$

Theorem 10.3.2 If $Y = aX + b$ then $M_Y(t) = e^{bt}M_X(at)$.

Proof.

$$M_Y(t) = E(e^{tY}) = E(e^{t(aX+b)})$$

$$= E(e^{atX}e^{tb}) = e^{tb}E(e^{(at)X})$$

$$= e^{bt}M_X(at). \quad (\text{ since } e^{tb} \text{ is a constant.})$$

Theorem:23.3.3 If all the higher moments $\mu_k = E(X^k)$ exist then
$M_X(k)(0) = \mu_k$,

that is, the k-th derivative at 0 of the m.g.f. of X equals the k-th moment.

Proof. Recall Taylor's formula from calculus (assume that it is convergent for all t here):

$f(t) = f(0) + f^1(0)t + f^{11}(0)$

We will apply this to $f(t) = M_X(t)$. Recall the power series expansion for e^x,

$e^x = 1 + x + x^2/2! + \ldots\ldots$

Replace x by tX to get

$e^{tX} = 1 + tX + t^2 X^2/2! + \ldots\ldots$ (1)

and take expectations on both sides, and use linearity of expectation and assuming that this is OK for an infinite convergent series:

$E(e^{tX}) = E(1) + E(tX) + E(t^2 X^2/2!) + \ldots\ldots$

$= 1 + tE(X) + t^2 E(X^2/2!) + \ldots\ldots$

$M_X(t) = 1 + t\mu_1 + t^2/2!\, \mu_2 + \ldots\ldots$ (2)

Now (1) applied to $f(t) = M_X(t)$ says

$M_X(t) = 1 + t M'(0) + t^2/2!\, M''(0) + \ldots\ldots$ (3)

so matching the coefficients in (2) and (3) gives the theorem.

This theorem can be used to easily calculate means, variances and higher moments.

Moment generating functions have many important and useful properties. One of the most important of these is the uniqueness property. That is, the moment generating function of a random variable is unique when it exists, so if we have two random variables X and Y, say, with moment generating functions $M_X(t)$ and $M_Y(t)$, then if $M_X(t) \quad M_Y(t)$ for all values of t, both X and Y have the same probability distribution.

Theorem.23.3.4.(Uniqueness Theorem) Suppose that X and Y are random variables having moment generating functions $M_X(t)$ and $M_Y(t)$, respectively. Then X and Y have the same probability distribution if and only if $M_X(t) = M_Y(t)$ identically.

23.4. Conditional Expectation

Recall that

- If X and Y are continuous random variables, then the conditional density function of Y given $X = x$ is given by

$$f_{Y/X}(y/x) = f_{X,Y}(x,y)/f_X(x)$$

- If X and Y are discrete random variables, then the probability mass function Y given $X = x$ is given by

$$p_{Y/X}(y/x) = p_{X,Y}(x,y)/p_X(x)$$

The conditional expectation of Y given $X = x$ is defined by

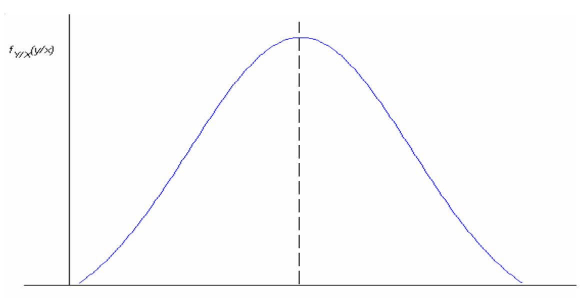
$$\mu_{Y/X=x} = E(Y/X=x) = \begin{cases} \int_{-\infty}^{\infty} y f_{Y/X}(y/x) & \text{if } X \text{ and } Y \text{ are continuous} \\ \sum_{y \in R_Y} y p_{Y/X}(y/x) & \text{if } X \text{ and } Y \text{ are discrete} \end{cases}$$

The conditional expectation of Y given $X = x$ is also called the conditional mean of Y given $X = x$. Clearly, $\mu_{Y/X=x}$ denotes the centre of mass of the conditional pdf or the conditional pmf as shown in Fig. below.

Remark

- We can similarly define the conditional expectation of X given $Y = y$, denoted by $E(X/Y = y)$
- Higher-order conditional moments can be defined in a similar manner.
- Particularly, the conditional variance Y given X = x of is given by

$$\sigma^2_{Y/X=x} = E[(Y - \mu_{Y/X=x})^2 / X = x]$$



Example:23.4.1.

Consider the discrete random variables X and Y discussed in example .The joint probability mass function of the random variables are tabulated in Table . Find the joint expectation of $\text{E}\ (Y / X = 2)$

x / y	0	1	2	$p_Y(y)$
0	0.25	0.1	0.15	0.5
1	0.14	0.35	0.01	0.5
$p_X(x)$	0.39	0.45	0.16	

The conditional probability mass function is given by

$$p_{Y/X}(y/2) = \ \mathrm{p}_{X,Y}(2, y) / p_X(2)$$
$$\therefore p_{Y/X}(0/2) = \ \mathrm{p}_{X,Y}(2,0) / p_X(2)$$
$$= \frac{0.15}{0.16} = 15/16$$

and

$$p_{Y/X}(1/2) = \ \mathrm{p}_{X,Y}(2,1) / p_X(2)$$
$$= \frac{0.01}{0.16} = 1/16$$
$$E(Y / X = 2) = 0 \times p_{Y/X}(0/2) + 1 \times p_{Y/X}(1/2)$$
$$= 1/16$$

Similarly, we can show that

$$p_{Y/X}(0/1) = \ 2/9$$

and

$$p_{Y/X}(1/1) = \ 7/9$$

so that

$$E(Y / X = 1) = 7/9$$

$p_{Y/X}(0/0) = 25/39$
and
$p_{Y/X}(1/0) = 14/39$
so that
$E(Y/X=0) = 14/39$

We also note that

$$EX = 0 \times p_X(0) + 1 \times p_X(1) + 2 \times p_X(2) = 0.77$$

and

$$EY = 0 \times p_Y(0) + 1 \times p_Y(1) = 0.5$$

Example 23.4.2.

Suppose X and Y are jointly uniform random variables with the joint probability density function given by

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2} & x \geq 0, y \geq 0, x+y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Find $E(Y/X=x)$

We have

$$\begin{aligned} \therefore f_X(x) &= \int_0^{2-x} f_{X,Y}(x,y)dy \\ &= \int_0^{2-x} \frac{1}{2} dy \\ &= \frac{1}{2}(2-x) \quad 0 \leq \text{x} \leq 2 \end{aligned}$$

$$\therefore f_{Y/X}(y/x) = f_{X,Y}(x,y)/f_X(x)$$

$$= \frac{1}{2-x}$$

$$\therefore E(Y/X=x) = \int_{-\infty}^{\infty} y f_{Y/X}(y/x)dy$$

$$= \int_{0}^{2-x} y \frac{1}{2-x} dy$$

$$= \frac{2-x}{2}$$

Example 23.4.3.

Suppose X and Y are jointly Gaussian random variables with the joint probability density function given by

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{X,Y}^2}} e^{-\frac{1}{2(1-\rho_{XY}^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho_{XY}\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]}.$$

Find $E(Y/X=x)$.

which is a Gaussian distribution.

Therefore,

$$E(Y/X=x) = \int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho_{XY}^2}} e^{-\frac{1}{2\sigma_Y^2(1-\rho_{XY}^2)}\left[(y-\mu_Y)-\frac{\sigma_Y\rho_{XY}}{\sigma_X}(x-\mu_x)\right]^2} dy$$

$$= \mu_Y + \frac{\sigma_Y\rho_{XY}}{\sigma_X}(x-\mu_x)$$

and

$$\text{var}(Y/X=x) = \sigma_Y^2(1-\rho_{XY}^2)$$

We can similarly show that

$$E(X/Y=y)=\mu_{X}+\frac{\sigma_Y \rho_{XY}}{\sigma_Y}(y-\mu_Y)$$

and

$$\operatorname{var}(X/Y=y)=\sigma_X^2(1-\rho_{XY}^2)$$

23.5. Conditional Expectation as a random variable

Note that $E(Y/X=x)$ is a function of x.

Using this function, we may define a random variable ϕ (X) = E(Y / X). Thus we may consider as a function of the random variable X and $E(Y/X=x)$ as the value of $E(Y/X)$ at $X=x$.

Theorem 23.5.1.

We establish the following results.

E[E[Y/X] = E[Y]

and

E[E(X/Y)] = E[X]

Proof:

$$\begin{aligned}
EE(Y/X) &= \int_{-\infty}^{\infty} E(Y/X=x) f_X(x)dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{Y/X}(y/x)dy f_X(x)dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_X(x) f_{Y/X}(y/x)dydx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)dydx \\
&= \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy \\
&= \int_{-\infty}^{\infty} y f_Y(y)dy \\
&= EY
\end{aligned}$$

Thus $E[E[Y/X] = E[Y]$ and similarly

$E[E(X/Y)] = E[X]$

The above results simplify the calculation of the unconditional expectations EX and EY. We can also show that

$E[E(g(Y)/X)] = E[g(Y)]$ $E[X]$

and

$E[E(g(X)/Y)] = E[g(X)]$

Example 10.5.1. Consider

$$EE(Y/X) = p_X(0)E(Y/X=0) + p_X(1)E(Y/X=1) + p_X(2)E(Y/X=2)$$

$E[E(Y/X)]$ $= 0.5$

$= EY$

23.6 Conditio$E[Y]$Variance

23.6.1 Marginal Variance: The **definition** of the (population) (marginal) variance of a random variable Y is

Var(Y) = E([Y - E(Y)]2)

There is another **formula** for Var(Y) that is sometimes useful in computing variances or proving things about them. It can be obtained by multiplying out the squared expression in the definition:

$\mathrm{Var}(Y) = E([Y - E(Y)]^2) = E(Y^2 - 2YE(Y) + [E(Y)]^2)$

Conditional Variance: Similarly, if we are considering a conditional distribution Y|X, we define the **conditional variance**

$\mathrm{Var}(Y|X) = E([Y - E(Y|X)]^2 \mid X)$

(Note that **both** expected values here are conditional expected values.)

Conditional Variance as a Random Variable:

As with E(Y|X), we can consider Var(Y|X) as a random variable. For example, if Y =

height and X = sex for persons in a certain population, then Var(height | sex) is the variable which assigns to each person in the population the variance of height for that person's sex.

Expected Value of the Conditional Variance:

Since Var(Y|X) is a random variable, we can talk about its expected value. Using the formula

$Var(Y|X) = E(Y^2|X) - [E(Y|X)]^2$,

we have

$E(Var(Y|X)) = E(E(Y^2|X)) - E([E(Y|X)]^2)$

We have already seen that the expected value of the conditional expectation of a random variable is the expected value of the original random variable, so applying this to Y^2 gives

(*) $E(Var(Y|X)) = E(Y^2) - E([E(Y|X)]^2)$

Variance of the Conditional Expected Value:

For what comes next, we will need to consider the variance of the conditional expected value. Using the second formula for variance, we have

$Var(E(Y|X)) = ([E(Y|X)]2) - [E(E(Y|X))]2$

Since $E(E(Y|X)) = E(Y)$, this gives (**)$Var(E(Y|X)) = E([E(Y|X)]^2) - [E(Y)]^2$.

Putting It Together:

Note that (*) and (**) both contain the term $E([E(Y|X)]2)$, but with opposite signs. So adding them gives:

$E(Var(Y|X)) + Var(E(Y|X)) = E(Y2) - [E(Y)]2$, which is just Var(Y). In other words,

(***) $Var(Y) = E(Var(Y|X)) + Var(E(Y|X))$.

In words: The marginal variance is the sum of the expected value of the conditional variance and the variance of the conditional means.

Consequences:

I) This says that two things contribute to the marginal (overall) variance: the expected value of the conditional variance, and the variance of the conditional means.

Moreover, $Var(Y) = E(Var(Y|X))$ if and only if $Var(E(Y|X)) = 0$.

II) Since variances are always non-negative, (***) implies

$Var(Y) =E(Var(Y|X))$.

III) Since $Var(Y|X) = 0$, $E(Var(Y|X))$ must also be = 0. Thus (***) implies

$Var(Y) = Var(E(Y|X))$.

Moreover, $Var(Y) = Var(E(Y|X))$ if and only if $E(Var(Y|X)) = 0$.

IV) Another perspective on (***)

i) $E(Var(Y|X)$ is a weighted average of $Var(Y|X)$

ii) $Var(E(Y|X) = E([E(Y|X) - E(E(Y|X))]^2)$

$= E([E(Y|X) - (E(Y)]^2)$,

which is a weighted average of $[E(Y|X) - (E(Y)]^2$

Thus, (***) says that Var(Y) is a weighted mean of Var(Y|X) plus a weighted mean of $[E(Y|X) - (E(Y)]^2$ (and is a weighted mean of Var(Y|X) if and only if all conditional expected values E(Y|X) are equal to the marginal expected value E(Y).)

23.7 Lesson End Excercises

1. Let the r.v.'s X, Y be jointly distributed with p.d.f. given by $f(x,y)= 2/((n(n+1))$ if $y = 1, \ldots, x$; $x= 1, \ldots, n$, and 0 otherwise. Compute the following quantities: $E(X|Y = y)$, $E(Y|X = x)$.

2. Let **X, Y** be r.v.'s with p.d.f. f given by f **(x, y)** = λ^2e-λ**(x+y)I**(0,8)×(0,8) **(x, y)**.

   Calculate the following quantities: **E[X],** σ^2**(X)**, **EY**, λ^2**(Y)**, **E(X|Y= y),** λ^2**(X|Y = y).**

23.8 Suggested Readings:

1. Goon,Gupta;Das Gupta (1991).Fundamental of Statistics
2. Gupta,S.C. and Kapoor,V.K. Fundamental of Mathematical Statistics
3. Hoel,P.G. (1971).Introductory of Mathematical Statistics
4. Hogg R.V and Craig,A.T.Introduction to Mathematical Statistics
5. Hogg ,RV and Tanis,EA(1993).Probability and Statistical Inference
6. Mood,AM,Bose DC and Graybill F A.Introduction to the Theory of Statistics
7. Rohtagi,VK.An Introduction to Probability Theory and Mathematical Statistics

****